Characterising Urban Space from Topographic Databases: Cartographic Pattern Recognition Based on Semantic Modelling

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Patrick Lüscher

von
Muhen AG

Promotionskomitee

Prof. Dr. Robert Weibel (Vorsitz, Leitung der Dissertation)
Prof. Dr. Dirk Burghardt
Prof. Dr. Werner Kuhn

Zürich, 2011

Summary

National Map Agencies and other data producers capture large volumes of topographic data at high detail. As these topographic databases were designed for a broad range of applications, they model geographic reality in terms of single objects such as houses, streets, and lawns. However, many applications require specific higher order geographic phenomena that are not available in these general purpose databases, such as the extent of a city centre. Hence methods are needed to abstract these higher order geographic phenomena from the detailed representations of topographic databases. The ambition of this research is to automate abstractions from the detailed concepts offered by topographic datasets to produce higher order geographic phenomena by means of cartographic pattern recognition.

Up to now, cartographic pattern recognition was mainly employed for optimising visual appearance. It can be argued, however, that it is primarily a task of modelling geographical content. The rationale of this research is to develop and evaluate an approach to cartographic pattern recognition that explicitly models semantic assumptions behind higher order geographic phenomena. With respect to this overall rationale, the research pursues the following three objectives: 1) Methods for knowledge acquisition to inform cartographic pattern recognition shall be explored; 2) instruments to model knowledge and compile the data enrichment process from semantically rich descriptions shall be developed; and 3) the role of uncertainty in the proposed approach shall be investigated. The objectives are pursued by means of two case studies: The first case study builds a typology of urban residential house types, formalises English terraced houses in a conceptual model and develops a method to transform the conceptual model directly into a pattern recognition process. The second case study focuses on city centres as an instance of a concept with vague definition and extent. Each case study performs the complete process from knowledge acquisition to evaluation of derived referents for higher level phenomena.

The main contribution of the research is a methodology to capture semantics of geographical phenomena in conceptual models and use it to execute cartographic pattern recognition. The second contribution is a method to integrate ontological modelling with Bayesian inference to carry out pattern recognition. The method combines structural knowledge with machine learning to overcome difficulties with the vague nature of terms that describe geographic phenomena. A third contribution is the application of participant experiments to acquire knowledge for cartographic pattern recognition.

Two main directions are seen to extend the research presented in this thesis towards an operational system. Firstly, workflow management systems could be integrated to allow efficient, yet flexible execution of complete pattern recognition workflows. Secondly, a comprehensive and operational system for cartographic pattern recognition would require appropriate human user interaction schemes and storage of relations.

Zusammenfassung

Topographische Datenbanken sind heutzutage in grosser Detailtreue verfügbar. Da diese Datenbanken jedoch für ein breites Spektrum von Anwendungen entworfen wurden, modellieren sie die geographische Realität sehr allgemein in Form von Einzelobjekten wie Häusern, Strassen und Grünflächen. Viele Aufgaben benötigen aber spezifische geographische Objekte mit komplexer Semantik, wie etwa die Ausdehnung des Stadtzentrums. Daher werden Methoden benötigt, um von den detailgetreuen Darstellungen der topographischen Datenbanken komplexe geographische Phänomene zu abstrahieren. Das Bestreben dieser Arbeit ist es, solche Abstraktionen mittels Methoden der kartographischen Mustererkennung zu automatisieren.

Bisher wurde kartographische Mustererkennung hauptsächlich eingesetzt, um die Darstellungsqualität während des Kartengeneralisierungsprozesses zu gewährleisten. Geometrische Operationen standen im Vordergrund, während die Semantik von geographischen Objekten wenig Beachtung fand. Die vorliegende Arbeit entwickelt und beurteilt eine Methode für kartographische Mustererkennung, die explizit die semantischen Annahmen berücksichtigt, die geographischen Phänomenen zugrunde liegen. In Bezug auf dieses übergeordnete Thema werden drei Zielsetzungen verfolgt: 1) Methoden, um Wissen für kartographische Mustererkennung zu erlangen werden untersucht; 2) Werkzeuge, um Wissen zu modellieren und den Mustererkennungsprozess von diesen Modellen abzuleiten, sollen entwickelt werden; und 3) die in geographischen Phänomenen inhärenten Unsicherheiten sollen berücksichtigt werden. Diese Zielsetzungen werden in zwei Fallstudien untersucht: Die erste Fallstudie erstellt eine Typologie von englischen städtischen Wohnhäusern, formalisiert den Wohnhaustyp *English terraced house* in einem konzeptuellen Modell und entwickelt eine Methode, um das konzeptuelle Modell direkt in einen Mustererkennungsprozess zu übertragen. Die zweite Fallstudie widmet sich dem

Stadtzentrum als Beispiel eines geographischen Konzepts mit vager Definition und Ausdehnung. Beide Fallstudien führen den gesamten Prozess von der Wissensaneignung bis zur Evaluation der durch die Mustererkennung abgeleiteten Objekte durch.

Der massgebliche Beitrag dieser Arbeit ist eine Methodik, um die Semantik geographischer Phänomene in konzeptuellen Modellen zu erfassen und dieses Wissen für die kartographische Mustererkennung zu nutzen. Der zweite wesentliche Beitrag ist ein Ansatz für die kartographische Mustererkennung, der ontologische Modellierung mit Bayes-Inferenz koppelt. Der Ansatz kombiniert strukturelles Wissen mit maschinellem Lernen, um vagen Begriffe zu handhaben, wie sie typischerweise in Beschreibungen von geographischen Phänomenen vorkommen. Ein dritter Beitrag ist die Anwendung von Nutzerbefragungen, um Wissen für die kartographische Mustererkennung zu generieren.

Schliesslich werden Wege vorgeschlagen, um die in dieser Arbeit dargelegte Forschung in Richtung eines operationellen Betriebs zu erweitern. Erstens sollen Systeme für *workflow management* integriert werden, um komplette Mustererkennungsabläufe transparent und flexibel modellieren und ausführen zu können. Zweitens müssen für die Realisierung eines operationellen Systems geeignete Ansätze für Benutzerinteraktionen und für die Speicherung der erzeugten höherwertigen Objekte gefunden werden.

Acknowledgements

This is the time and place to express my sincere gratitude to all who, in one way or another, have helped me produce this thesis.

First and foremost I thank my supervisors Robert Weibel and Dirk Burghardt, who gave me—after having supervised my Diploma thesis already—the opportunity to carry out this research in the GIS unit. I am deeply indebted to Rob and Dirk for their support and patience during the time of my PhD, for proof-reading papers and this thesis, and for their valuable comments.

I am most grateful to Prof. Dr. Werner Kuhn for being a member of my PhD committee. I would like to thank Prof. Dr. Lars Harrie for acting as an independent reviewer of this dissertation.

Nicolas Regnauld of Ordnance Survey Research made it possible that the Ordnance Survey of Great Britain funded part of my research and provided me with data. I am deeply indebted for his support. Many thanks to the generalisation team in Ordnance Survey Research for taking time for discussions.

William Mackaness was willing to collaborate with me during a three-week visit at the University of Edinburgh in November 2007. William impressed me with his imaginativeness and his enthusiasm for research and helped sharpen my ideas. In the bespoke visit in Edinburgh, I also have become acquainted with Omair Chaudhry, who was then William's PhD student. Coincidentally, he was later research associate at Ordnance Survey at the time of my collaboration. I would like to thank William and Omair for our conversations.

Numerous people at the Department of Geography have made my time as PhD student an inspiring and enjoyable one. I am indebted to Ross Purves and Sara Fabrikant for running the promotion seminar and teaching me how to do research. Despite his many commitments, Ross always took the time and helped me with critical feedback about my work. I am thankful to my fellows for interesting conversations about scientific and non-scientific issues, fun times, consoling words, and just for being friends: Pia Bereuter, Somayeh Dodge, Alistair Edwardes, Anna-Katharina Lautenschütz, Moritz Neun, Ronald Schmidt, Stefan Steiniger, Ralph Straumann, Martin Tomko and Ramya Venkateswaran.

I would like to thank Elisabeth Cottier, who ran the GIS unit's secretariat for many years, and Annica Mandola, who then replaced Elisabeth, for their support in all matters of administration and for their interest in us PhD students.

I am most grateful to my parents for supporting my studies and motivating me to take the next step.

And finally, I would like to thank Sonja for always being there, for her patience and understanding during the time of my dissertation.

This research was supported by the Swiss State Secretariat for Education and Research (SER) through COST Action C21 (project ORUS, Grant no. C05.0081). I am grateful to the Ordnance Survey of Great Britain for funding part of the research and for provision of data.

Contents

List of Figures

Chapter 3

Chapter 4

Appendices

Part I

Synopsis

Chapter 1

Introduction

1.1 Motivation

1.1.1 General purpose databases, very specific needs

Traditionally, topographic information was produced and disseminated by National Mapping Agencies (NMAs) in the form of paper maps that were general purpose, high-quality products. This genericity is reflected in current topographic datasets. These datasets offer a wealth of accurate (but mainly geometric) information about individual objects. However, they do not model specific higher order semantics required by many applications. For example, they represent individual houses, parking spaces and ponds, but not residential areas, city centres and mountain ranges. Hence, current topographic datasets need to be enriched with additional geographic meaning (Thomson, 2009). The tenet of this thesis is that semantic enrichment of topographic maps can be automated by means of cartographic pattern recognition. This firstly helps NMAs and other data producers to establish a more user-driven access to geographic information and respond better to varying user needs (Hart & Greenwood, 2003; Davies, Wood, & Fountain, 2005), as different clients have specific needs for certain representations. Secondly, cartographic pattern recognition allows better adapting representations to the way how people conceptualise human space. Overcoming the semantic mismatch between available models and actual use of geographic information is a long-standing research topic in geographic information science (Couclelis, 2009; Schuurman, 2006; McMaster & Usery, 2004; Müller et al., 1995).

1.1.2 General purpose databases, many applications

Cartographic pattern recognition for semantic enrichment helps in a wide variety of application areas. Government agencies integrate a wealth of geographic information for analysis and decision making, in particular for many forms of analysis in *urban planning*. For example, building character and age are important parameters to estimate energy efficiency (Jones et al., 2007). Automated methods for building classification can be used to map such parameters on a large scale. Automatic discovery of urban functional units such as a city centre, shopping districts and amusement districts can be used to monitor the functional structure and evolution of urban areas and can hence be beneficial for planning and monitoring urban regeneration (Bromley et al., 2003; Tallon & Bromley, 2004).

Another application area for which cartographic pattern recognition is useful is *interoperability of spatial information*. Since different datasets commonly rely upon different conceptualisations of the world, showing different user views and being captured at different levels of abstraction, their information content must be harmonised before integration can happen (Bishr, 1998). In different datasets, semantically equal phenomena can be named differently, or unequal phenomena can have the same name. Pattern recognition based on semantic modelling can be used to resolve such ambiguities (Klien, 2008).

Automated techniques for cartographic pattern recognition were probably first called for by the map generalisation community. Here, cartographic pattern recognition is needed to *build and update multiple representation databases.* MRDBs store geographic phenomena at multiple conceptual levels (Kilpeläinen, 1997). There are two approaches for creating MRDBs. The first approach is through the integration of existing representations, which requires data matching techniques. The second approach uses derivation from a base representation, which requires map generalisation operations. The latter approach is favourable due to lower cost for capture and update. Hence, approaches to automate abstraction of datasets are sought.

Finally, human spatial reasoning is chiefly qualitative, i.e. based on spatial relations and regions (Egenhofer & Mark, 1995; Montello, 2003). Representing geographic regions is thus beneficiary for applications such as g*eographic information retrieval*. Queries posed on web search engines such as Google and Microsoft Bing often contain a spatial component (Purves et al., 2007). For example, people might search for “shopping opportunities in the city centre of Zurich”, or “cafés in the old town”. Thus, cartographic pattern recognition can be used to create higher order phenomena that form the context of such queries (Purves et al., 2007; Vögele et al., 2003; Larson, 1996). Representing such regions would also help to provide a better sense of place in mobile information systems (Tamminen et al., 2004).

1.1.3 An example: Semantic modelling for automated map generalisation

In the context of topographic maps, interpretation of geographic information into more abstract form is termed map generalisation. The International Cartographic Association defines map generalisation as *"selection and simplified representation of detail appropriate to the scale and/or the purpose of a map"* (ICA 1973, p. 173). In a digital context, two kinds of generalising spatial information can be distinguished: *Model generalisation* includes various operations to transform a spatial database, without aiming at visual presentation; processes envisaged to optimise geometric quality for visualisation are subsumed as *cartographic generalisation*.

The broad definition above encompasses many forms of transformation. However, a large part of research on automated map generalisation has focused on reproduction of the content of traditional topographic maps, and on handling small transitions of scale, where there are little changes in conceptualisation of the represented phenomena.

The ongoing struggle to achieve more drastic abstractions has been attributed to a failure to envisage the generalisation process as a task of modelling geographic meaning (Nyerges, 1991; Harvey, 1997; Mackaness, 2006), rather than *"something you do at the end"* to polish the aesthetic appearance of a map (Mackaness, 2007). As interpretation of spatial information is a knowledge-intensive task (Minsky, 1975), methods for knowledge acquisition and representation gained attention in research on automated map generalisation (Weibel et al., 1995). The term *structural knowledge* refers to the domain knowledge involved in the map generalisation process (Armstrong, 1991). The generalisation community promoted a phenomenological perspective that explores *"how geographic phenomena merge or separate to create higher order, more generalized forms"* (Chaudhry, 2007, p. 9). Mackaness (2006, p. 251) reports on a quote attributed to Minsky that says: *"you cannot tell you are on an island by looking at the pebbles on the beach"*. However, as Mackaness argues, it is made possible if we understand what an island is and how to search for it within a database of pebbles. Lüscher et al. (2007) and Thomson (2009) discussed the potential of semantic modelling to provide this additional synoptic view.

1.1.4 Focus on urban structure

The world faces an ongoing process of urban growth. The proportion of urban population in Switzerland has reached 73% in the year 2000 and the land covered by urban areas has doubled between 1980 and 2000 (Bundesamt für Raumentwicklung, 2009); it is supposed that this urbanisation process has not yet finished (Baccini et al., 2007). On the one hand, daily experience of geographic space is nowadays, to a major part, affected by urban landscape. On the other hand, the continuing sprawl of urban areas keeps issues of urban liveability and sustainability as major issues on the agenda of research and public policy making.

The spatial pattern of cities affects physical, ecological, and socioeconomic processes within their boundaries and beyond (Luck & Wu, 2002). Urban morphology examines the structure of the city, and the role of humans that organise it, through analysis of its shape. However, a major hurdle to large-scale analysis of urban form is the poor availability of appropriate data (Lannon & Linowski, 2009). In this respect, the potential of geographic information is not yet fully exploited (Thomson, 2009).

1.2 Thesis rationale

As was argued above, making higher order geographic phenomena explicit in topographic datasets is seen critical to render them more versatile, better accessible, and more intelligent. This thesis approaches adaptation of topographic databases to specific needs by means of cartographic pattern recognition and adopts a phenomenological view. A main issue in this context is how to acquire and model phenomenological knowledge so that it can be exploited to guide the model generalisation process. As will be seen, this touches upon issues of ontology and human conceptualisation. The following sections describe the research objectives in more detail, and relate research questions to each objective.

1.2.1 Research objectives and methodology

The main objective of this thesis is to develop a methodology to enrich spatial datasets. The enrichment process is to be based upon semantically rich descriptions of geographic phenomena. The targeted methodology is termed *ontology-driven data enrichment* (Paper 1).

The approach adopted within the methodology consists of three basic steps (Figure 1.1). *Knowledge acquisition* refers to methods for obtaining semantically rich conceptual models of urban structures. *Knowledge formalisation* refers to methods of describing the acquired knowledge. *Data enrichment*, finally, refers to the execution of the pattern recognition process.

Figure 1.1: Basic steps of the ontology-driven approach

In line with the main objective of this thesis and the requirements analysis performed in Paper 1, three research objectives are formulated that structure the problem solving process of this work:

Research objective 1: *The research shall explore methods for semantic grounding.* Ontology-driven data enrichment is to be based on semantically meaningful concepts. However, it is unrealistic that a single catalogue of urban structures can be built that satisfies all possible applications. Firstly, urban processes are dependent on cultural influences, and hence is the genesis of urban structures. Secondly, definitions and interpretations of the same urban structure may vary considerably amongst different groups of users, while each definition exists in its own right. The challenge is therefore to identify standard approaches for acquiring knowledge and modelling for data enrichment.

Research objective 2: *The research shall develop instruments to model knowledge and derive the data enrichment process from semantically rich descriptions.* The acquired knowledge of urban structures has to be described in formalised models. Finally, the models have to be related to spatial analysis operations in order to enrich the database. The ambition is to develop a procedure that is based on generic modules and minimises implementation efforts. Previous works proposed to employ description logic reasoning (Thomson, 2009) or rule-based reasoning (Klien, 2007) for achieving both goals. The practicability of such approaches for large datasets shall be investigated.

Research objective 3: *The research shall investigate the role of uncertainty in the ontology-driven data enrichment approach.* Uncertainty manifests itself in a number of ways in geographic phenomena (see Section 2.2). Whereas discord (uncertainty from varying conceptualisations; Fischer, 1999) has received considerable attention, there is a lack of research on methods for representing vagueness for data enrichment.

Note that in a complete database enrichment process, the last step would be to transfer the enhanced knowledge back to the database. By this procedure, a multiple representation database (MRDB; Kilpeläinen, 1997) is created. However, the creation of MRDBs is not explicitly covered in this thesis.

1.2.2 Research questions

Six research questions are formulated to be answered in this thesis. Research questions I–II relate to the overall methodology. Research question III relates to issues of grounding. Research questions IV–V relate to pattern formalisation within the ontology-driven methodology. Finally, research question VI relates to issues of uncertainty.

(I) How can semantic modelling help in the development of cartographic pattern recognition methods?

(II) What are the requirements for an ontology-driven approach to data enrichment in an urban context?

(III) What methods are available for extracting knowledge about urban structures?

(IV) Can urban structures be decomposed in terms of the phenomenological approach?

(V) To what extent is it possible to use only simple measures (such as area and topological relations) to define complex concepts?

(VI) How can we integrate vagueness into the data enrichment process?

1.2.3 Research papers

This thesis is based on four successively published research papers. All papers underwent peer reviews based on full manuscripts. Figure 1.2 provides a graphical summary of the contributions of each paper to the objectives set out in Section 1.2.1. The four research papers are:

Research paper 1:

Lüscher, P., Burghardt, D., & Weibel, R. (2007). Ontology-driven Enrichment of Spatial Databases. *10th ICA Workshop on Generalisation and Multiple Representation*, Moscow, Russia, August 2–3, 2007.

Research paper 2:

Lüscher, P., Weibel, R., & Mackaness, W. (2008). Where is the Terraced House? On The Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases. In A. Ruas & C. Gold (Eds.), *Headway in Spatial Data Handling. Proceedings of the 13th International Symposium on Spatial Data Handling* (pp. 449–466). Berlin / Heidelberg: Springer-Verlag.

Research Paper 3:

Lüscher, P., Weibel, R., & Burghardt, D. (2009). Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. *Computers, Environment and Urban Systems, 33*(5), 363–374.

Research Paper 4:

Lüscher, P., Weibel, R. (submitted). Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases. *Computers, Environment and Urban Systems*, revised manuscript submitted June 2011.

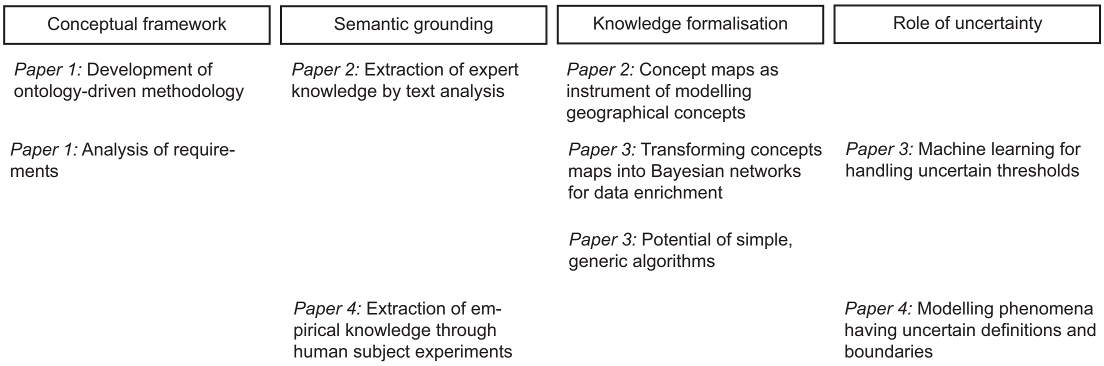


Figure 1.2: Contributions of each research paper to the individual research objectives

Paper 1 develops the ontology-driven methodology. It analyses shortcomings of purely algorithmic approaches to cartographic pattern recognition and specifies requirements for an ontology-driven approach. Papers 2 & 4 are dedicated to methods for acquiring knowledge about urban concepts. Paper 3 deals with possibilities of standardising the pattern recognition workflow in the light of uncertain knowledge of thresholds. Paper 4 shows how knowledge about vaguely defined urban structures can be acquired through human subject experiments, and suggests approaches for verification of model outputs. The ontology-driven approach is assessed based on two case studies. The first case study (Papers 2 & 3) examines residential building classification. The second case study (Paper 4) examines locating and delineating a city centre. Hence, the second case study is more complex because the conceptualisation of a city centre is inherently vague, and because a city centre cannot be created through aggregation operations from lower level concepts only.

1.3 Structure of the Thesis

This thesis is organised into two parts. Part I (Synopsis) integrates the above research papers in the scientific context. Part II (Research Papers) presents the research papers with the content and format as they were submitted or published. Additionally, Part III (Appendices) provides background information on the datasets that were used for this study and contains the full list of publications that were created in the course of this research. The structure of the Synopsis is as follows:

Chapter 1 *Introduction.* The current chapter presents the motivation and develops the rationale for the research. The objectives and research questions are defined.

Chapter 2 *Background and State of the Art.* The second chapter provides the reader with information that forms the background and context of this research, reviews the State of the Art in pattern recognition for map generalisation, and concludes with stating challenges for research.

Chapter 3 *Summary of Papers.* The third chapter presents an executive summary of the four publications. Each paper is summarised in terms of its rationales, methods, results, and contributions.

Chapter 4 *Discussion.* The fourth chapter revisits the research questions set out in the current chapter and provides an integrated discussion of the results of the research.

Chapter 5 *Conclusions and Outlook.* In the fifth chapter, the relevance and contributions of the research in the context of geographic information science is discussed. The Synopsis concludes with an outlook on potential future research.

Chapter 2

Background and State of the Art

The aim of this chapter is threefold. Firstly, the chapter introduces the topics that form the context of the research. Secondly, the State of the Art in cartographic pattern recognition is presented. Thirdly, research challenges with respect to this context are highlighted.

Section 2.1 explores the changing nature of geographical phenomena with respect to purpose, or scale, of a topographic database. The process of abstracting geographical information is known as map generalisation, and data enrichment by cartographic pattern recognition was devised as part of the solution for its automation.

Geographic phenomena are subject to various types of uncertainty. It is a tenet of this research that uncertainty has to be respected while carrying out spatial data enrichment. Section 2.2 provides the reader with the theoretical background about uncertainty of geographic phenomena.

The research was carried out in an urban context. Section 2.3 reports on concepts for analysing urban structure from an urban planning perspective.

Section 2.4 reports on existing approaches for data enrichment, focusing on enrichment of topographic vector databases, and on the urban environment.

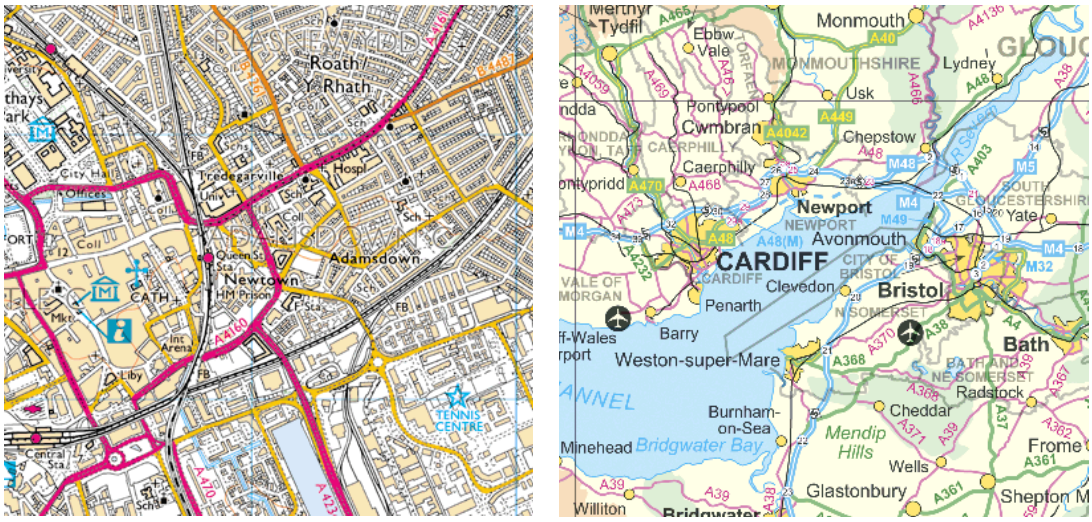
Finally, Section 2.5 states research challenges and lists the contributions of this research.

2.1 Generalisation of geographic phenomena

2.1.1 Scale, purpose, and map generalisation

Spatial database models are abstracted representations of a portion of the real world. The same reality can be abstracted in many different ways, subject to the purpose the model serves for (Haggett & Chorley, 1967). Purpose and scale are functionally related subjects. Many geographic phenomena have a scale at which they operate and hence it is vital to study them at the appropriate scale (McMaster & Sheppard, 2004). Likewise, scale-dependent meaning of concepts such as *place*, *neighbourhood*, and *region* is inherent in human thinking (Agarwal, 2004; Agarwal, 2005b; Montello, 1993; Klippel et al., 2009).

The map generalisation process can then be defined as deriving from a detailed model one that is usually higher abstracted, but better focused on a specific purpose (Figure 2.1). The aim of automated generalisation is to conduct this process with little or no manual interaction. The ultimate ambition of research on map generalisation processes is to be able to automatically derive representations for arbitrary scales and purposes from a single, highly detailed database (Weibel, 1997).



(a) (b)

Figure 2.1: (a) Exploring the urban structure and (b) relating settlements (you cannot travel from Cardiff to Bristol without crossing the water channel)—different tasks require specific representations. (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved.)

2.1.2 Model generalisation and cartographic generalisation

Generalisation in the digital domain involves modelling as well as representative tasks (Weibel & Dutton, 1999). Figure 2.2 shows the stages of a generalisation workflow (Grünreich, 1985). A basic distinction is made between object/model generalisation on the one hand, and cartographic generalisation on the other hand.

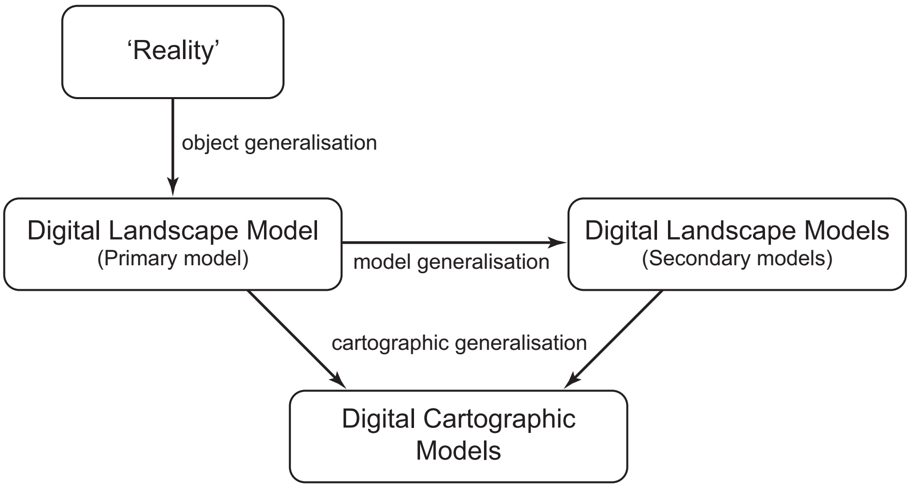


Figure 2.2: Generalisation as a sequence of modelling operations (after Grünreich, 1985)

Object generalisation happens at the time of defining and building the initial database. The diversity of reality is reduced subject to various reasons, such as intended use, sampling methods, and human interpretation skills (Weibel & Dutton, 1999). *Model generalisation* derives special-purpose secondary models from the primary model. It involves spatial analysis operations to alter and reduce data for various purposes. Finally, the landscape models produced by object and model generalisation are prepared for visualisation representation through *cartographic generalisation*. It deals with problems that are created by symbolisation, such as congestion of map features that impede good readability of a map. Hence, model generalisation might serve as an intermediate step before visualisation, but more importantly, it serves as a preparation step for many analytical applications (some examples were highlighted in Section 1.1).

2.1.3 Data enrichment for generalisation

Map generalisation tends to simplify by selecting and emphasising essential content, while suppressing and omitting unimportant elements (Weibel, 1997). The way individual objects are generalised depends on the spatial context (Mustière & Moulin, 2002). Relations with surrounding objects must be respected, such as whether an object is part of a significant group or within a certain area. Hence, in many generalisation frameworks a structure recognition step is introduced that makes spatial context explicit (Ruas & Plazanet, 1996; Brassel & Weibel, 1988).

The term *data enrichment* is generally used for the activity that provides supplementary information or improves the quality of the original database[1]. Application of data enrichment in generalisation ranges from computation of neighbourhood graphs to characterisation of urban neighbourhoods (Neun et al., 2008). In this thesis, the term data enrichment is used for deriving additional geographical higher level knowledge from a cartographic database by pattern recognition techniques.

2.1.4 Focus on phenomena

As a portion of the earth is increasingly more abstracted, there are moments when the conceptualisation fundamentally changes and higher order phenomena emerge, such as when *"a city emerges from a collection of houses and streets"*, or a *"coal pan from a collection of coal mines"* (Bertin 1967, 1999, p. 300). Muller (1991) terms these events 'catastrophic change'; Bertin uses the term *conceptual generalisation* for the process of creating higher order phenomena out of a collection of more basic phenomena.

The emergence of higher order phenomena is a major challenge in map generalisation research. A solution lies in seeing generalisation as a modelling problem (Mackaness, 2007; Mackaness & Edwards, 2002; Mustière et al., 2000), and by basing the generalisation process on the phenomenon being mapped (Ormsby & Mackaness, 1999). This has promoted research in *phenomenological generalisation* (Mark, 1989). Taking a phenomenological perspective means to explore *"how geographic phenomena merge or separate to create higher order, more generalized forms"* (Chaudhry 2007, p. 9).

Ormsy and Mackaness (1999) divide the composition of a phenomenon into aspects of *geometry*, *semantics*, and *inter-object relationships*. Semantics in spatial databases is expressed in the classification of objects (for instance, a house, a street, or a river). Considering not only two dimensional topographic data where geometry is commonly the only available property, geometry can be generalised to *qualities* (Figure 2.3). Qualities are perceivable or measurable properties of entities (Masolo et al., 2003). In addition to geometry, other qualities that are important in an urban setting are for instance building height, construction period, or type of facade.

[1] http://www.information-management.com/glossary/d.html. Accessed 31.03.2011.

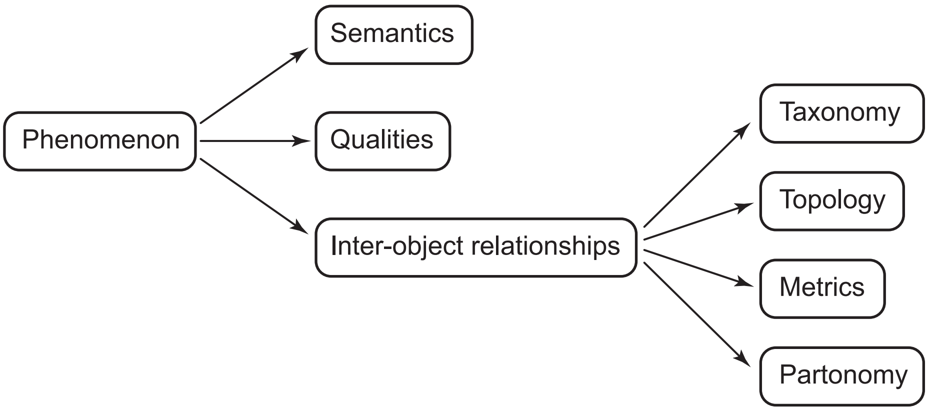


Figure 2.3: Aspects of modelling geographic phenomena in generalisation research (extended from Ormsby & Mackaness, 1999)

Van Smaalen (2003) distinguishes between the same three aspects, and details inter-object relationships further into *thematic* and *spatial relationships*. He only considered topological relations for the latter category. Chaudhry (2007) adds partonomic relations as a way of expressing membership with respect to higher order phenomenon. Since spatial relations such as proximity are equally important as topological for the constitution of many phenomena, Figure 2.3 adds *metrics* as another type of spatial relationship. Steiniger and Weibel (2007) identify two additional categories of relationships: *Statistical and density relationships*, and *structural relationships*. However, they are both combinations of the above listed, fundamental types. The types of inter-object relationships can be described as follows:

- The granularity of an object's classification can be different. For example house, garage, and factory are all kinds of building, which is again a kind of man-made structure (among roads, water pipes, and many other things). Thus, classifications form a hierarchical system. A *taxonomy* is a particular classification system, arranged in a hierarchical structure.
- *Topological relationships* are preserved under continuous transformations of space. Examples of topological relationships are disjoint, adjacent, contained in, etc. (Egenhofer & Herring, 1990).
- *Metrics* encompass distance and directional relationships, such as proximity and 'in front of'. Hence, their significance is similar to topological relationships. However, metric relationships vary under transformations of space.
- *Partonomy* relates objects with respect to a higher order phenomenon. It expresses that a collection of objects build a functional unit, forming together a higher level phenomenon. For example, a building, a yard, and the access way, are all part of the same higher

order phenomenon ‘lot of land’. The same principle applies for a collection of buildings, gardens and roads, forming together a settlement.

Modelling higher order phenomena in terms of the above discussed categories can be used to abstract from basic to higher level phenomena (Molenaar, 2004). Van Smaalen (1996, 2003) shows an example to build urban land use patches, departing from individual database objects (Figure 2.4). Liu et al. (2003) present a model for hierarchical aggregation of areal partitions where the partonomic information is expressed as a similarity matrix. Other examples of phenomenological modelling for higher order phenomena are discussed in Section 2.4.5.

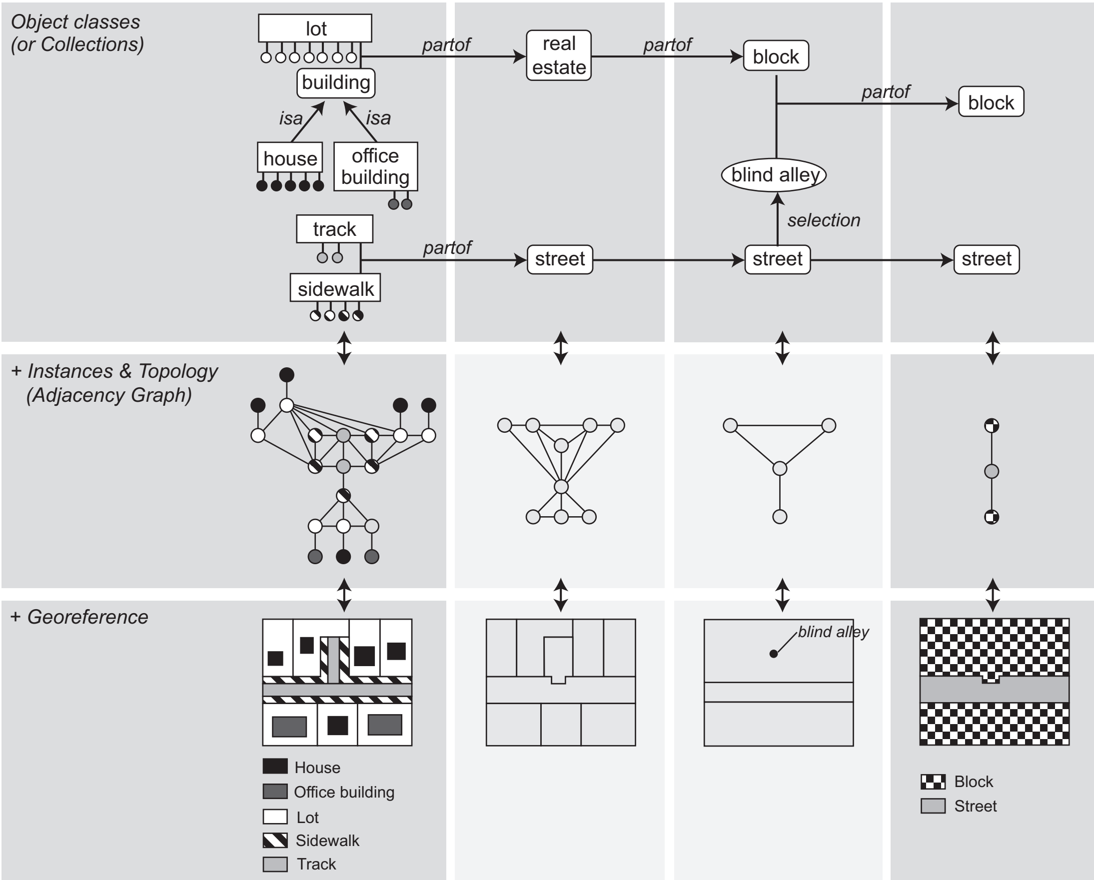


Figure 2.4: Functional object aggregations to urban land use patches (van Smaalen, 1996, p. 69)

2.2 Uncertainty of spatial information

2.2.1 Nature of uncertainty in spatial phenomena

Uncertainty is a quality of not being definitely known or knowable, or of being indeterminate as to magnitude or value (Simpson & Weiner, 1989). Spatial data can be subject to different types of uncertainty. It is important to distinguish between them in order to deal with each type of uncertainty properly. Fisher (1999) presents a conceptual model of uncertainty in spatial data (Figure 2.5). On the most basic level, Fisher distinguishes between uncertainty where class and/or instances[1] are well defined and uncertainty where class and/or instances are poorly defined. If there is no problem of separating the objects into clear-cut classes, then the phenomenon is said to be well defined and uncertainty is only due to error (i.e., imperfections in the measurement, or out-dated information). This type of uncertainty is common in most sciences.

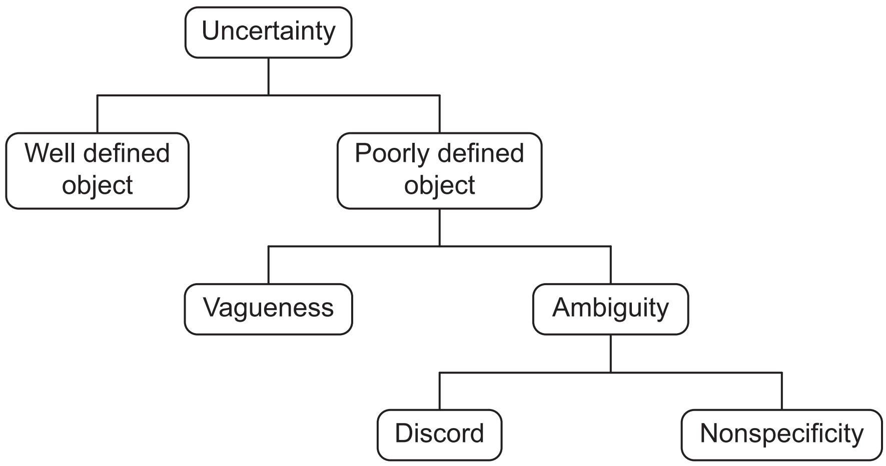


Figure 2.5: Conceptual model of uncertainty in spatial data (Fisher, 1999)

However, many spatial phenomena are uncertain in their definition. This type of uncertainty has its roots in philosophical and cognitive aspects of spatial information. It can be extensional, i.e. relating to assignment of objects to classes, or intensional, i.e. relating to descriptions of classes and class systems. Extensional uncertainty is termed *vagueness* by Fisher. Intensional uncertainty, termed *ambiguity* by Fisher, has again two subcategories. *Discord* in the case of the soil map means that there are many soil classification systems and the same patch might be assigned to a different soil type depending on the classification system one is using. *Nonspecificity* means that there is no equivocal set of conditions available for defining a phenomenon (Bennett, 2001).

A very similar taxonomy of uncertainty is also made by Bennett (2001), although he terms them *sorites vagueness* (instead of vagueness), *conceptual vagueness* (instead of nonspecificity), and *ambiguity* (instead of discord). This terminology emphasizes that vagueness and nonspecificity are closely related since vagueness is often caused by nonspecificity. Hence,

[1] A class in programming is defined as a set of objects with common properties. Instances are members of a class. The class ‚capital city‘, for example, has the instances London, Paris, Berlin, etc. The definition of a class is also called intension, and the set of members is called extension of a class.

in the following detailed discussion, vagueness and nonspecificity are treated in the same section, while discord is discussed separately.

2.2.2 Vagueness

According to Williamson (1994), a vague predicate is one that is susceptible to the Sorites Paradox. Originally, the Sorites Paradox was formulated as "how many grains of sand does it take to make a heap?" The Sorites argument is developed in the following way:

Premise	Fx_1	One grain does not make a heap.
Modus ponens	$(\forall_i)(Fx_i \rightarrow Fx_{i+1})$	Adding one grain to a 'not heap' does not turn it into a heap.
Conclusion	Fx_n	No matter how many grains are added, there is no heap.

Although the premise and the modus ponens seem plausible, the conclusion is obviously wrong. Many predicates (such as heap) do not have clear-cut boundaries, but there seems to be a gradual transition. At some point the judgment switches from 'not heap' to 'heap' for no obvious reason (Goldstein, 2000). Many geographical phenomena exhibit this kind of uncertainty: What is the difference between a hamlet and a village? And between a hill and a mountain? Where are the limits of a city (Fisher, 2000a)?

There are three stances that are debated concerning the nature of vagueness (Earl, 2010): The position that vagueness is an intrinsic property of phenomena themselves is termed *ontic vagueness*. *Epistemic vagueness* takes the stance that phenomena are of crisp nature, but that the exact boundary is not (or cannot be) known precisely. Finally, it can be argued that vagueness arises from individual interpretations of the world, each interpretation being crisp on its own. The last case is termed *semantic vagueness* (Varzi, 2001; Bennet, 2010).

Bennett (2010) further differentiates between vagueness of different linguistic categories. One the one hand, *attributes* such as 'large', 'steep', and 'tall' can exhibit vagueness. Bennett (2010) argues that vagueness of *noun predicates*, such as 'mountain', 'city', and 'lake', is generally more complex than that of attributes. While attribute vagueness is often dependent on one measure, the vagueness of a concept such as 'city extent' involves many different types of information, such as density of housing, distribution of retail and services etc. Finally, the third linguistic category is made up of *relations* such as 'near to' and 'north of'.

2.2.3 Categorisation and prototype theory

By category a number objects is meant that are considered equivalent (Rosch, 1987). Categorisation, i.e. attribution of a thing to a category, happens every time we see something as a *kind of* thing. It is seen as one of the most fundamental principles of human reasoning (Lakoff, 1987). The classical categorisation theory assumes a set-theoretic view on categories, i.e. every possible membership is either a member or not a member of each particular set. Furthermore, it is assumed that every member would be an equally good member of that set, and that there are rules available for determining set membership (Mark, 1993; Lakoff 1987, pp. 6–7; Smith & Mark, 1998).

It was later realised that classic categorisation theory is not entirely wrong, but only part of the story (Lakoff, 1987, p. 5). *Prototype theory* emerged to explain the complexity involved in categorisation (Rosch, 1978). Prototype theory established that categories may have an internal structure, since it was observed that people judge certain members of categories as being more representative of the category than others. For example, robins and sparrows are judged to be the best examples of the category 'bird', while owls and eagles are less so, and ostriches, emus, and penguins are among the worst examples (Lakoff, 1987, p. 44). Rosch (1978) termed the notion *prototype* for the most representative examples of a category and proposed to use *degree of prototypicality* as descriptor of categories. A second tenet of prototype theory is that while certain categories have clear boundaries, other categories have not only an internal structure, but also fuzzy boundaries.

Rosch's studies on categorisation were mainly based on "human-sized" objects, where "what" and "where" are most always independent (Smith & Mark, 1998; Lakoff, 1987, p. 51). In contrast, the "what" and "where" are intimately intertwined in the geographical world (Smith & Mark, 1998). However, Mark and Turk (2003, p. 30) state that *"empirical evidence appears to show that geographic categories have the same sorts of structures and internal organizations as do categories in other domains"*, even though graded boundaries are more common for geographic entities. Smith and Mark (1998) therefore distinguish between *fiat boundaries*, which correspond to genuine discontinuities in the world, and *bona fide boundaries*, which are projected onto the world by human cognition and language.

2.2.4 Dealing with vagueness of spatial information

Theories for dealing with vagueness in GIScience can be assigned to three groups, which will be discussed in the following.

2.2.4.1 Fuzzy sets

Fuzzy set theory (Zadeh, 1965) is probably the most prominent approach to handle vagueness in the GIScience literature. Examples are mapping of soil types (Burrough, 1989), land value evaluation (Sui 1992), integration of categorical maps (Hagen, 2003), and extraction of landscape features from digital terrain models (Fisher et al., 2004).

Fuzzy set theory is an extension of classical Boolean set theory. In classical set theory, the *law of excluded middle* dictates that each entity is either part of a set, or not (Williamson, 1994, p. 9). Fuzzy set theory abandons this assumption by defining a fuzzy membership function μ, $0 \leq \mu \leq 1$, which denotes the degree to which an entity is part of a set (Figure 2.6).

There is also an elaborate set of tools to support reasoning and decision making using fuzzy sets (cf. Robinson, 2003).

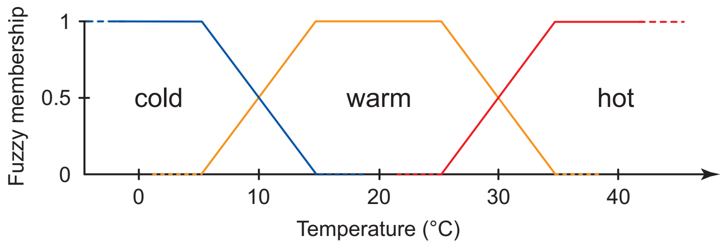


Figure 2.6: Example of fuzzy membership functions for air temperature terms

2.2.4.2 Supervaluation semantics

In the view of supervaluation semantics a vague predicate is one that allows several interpretations (termed *precifications*; Varzi, 2001). For example, the urban area of Bristol can be given a precise meaning by drawing a boundary line. There might be many precifications. There are precifications that are true in all interpretations (termed *super-true*) and precifications that are false in all interpretations (termed *super-false*).

In contrast to fuzzy logic, where each interpretation is a subset of a less rigid interpretation, supervaluation semantics does not impose that precifications are ordered. Hence, it is more generic than fuzzy logic. Another benefit of supervaluation semantics is that it allows keeping the instruments of classical logic for reasoning (Kulik, 2001).

Despite these benefits, applications of supervaluation semantics to geospatial problems are rare. Bennet (2001) investigates supervaluation semantics for defining forested areas. Santos et al. (2005) use supervaluation semantics for extraction of hydrographic features from maps.

2.2.4.3 Other means of representing vagueness

The ‘Egg-Yolk’ representation by Cohn and Gotts (1996) shares both properties from fuzzy sets and supervaluation semantics. Cohn and Gotts (1996) suggest to partition space into three regions with respect to membership to a phenomenon: The ‘yolk’, which is always part of the phenomenon, the ‘outside’, which is never part of the phenomenon, and the ‘white’, the remaining space where membership is contested. Thus, it allows keeping some of properties of classic logic, while being restricted to a concentric view of interpretations of the world.

Fuzzy sets can also be interpreted probabilistic. For instance, a value of $P(heap_i) = 0.9$ would mean that a certain amount of sand is denoted as heap in 90 % of the cases. Montello et al. (2003) discuss this stance for representing ‘downtown’.

2.2.5 Discord and ontologies

Discord arises because of different conceptualisations of the world. A conceptualisation is, according to Smith and Mark (2003, p. 414), *”a system of concepts or categories that divides up the pertinent domain into objects, qualities, relations, and so forth”*. For example, there are cross-cultural differences in the meaning of categories for standing water bodies (Mark, 1993). Fisher (1999) points out that there are many different soil classification systems and hence the same patch of land can be classified differently, depending on the classification system one is using. Often, there is no direct match of categories in different systems, but the categories overlap partly. The English term ’river’ overlaps with both French terms ‘fleuve’ and ‘rivière’ (for more examples see Mark, 1993).

Such kinds of ambiguity are a major impediment for information integration, interoperability of information systems, and for human-computer interaction (Smith & Mark, 1998). Hence, the study of the kinds of entities that make up the world, subsumed as *ontology*, has gained increased attention within geographical information science. Understanding and use of ontology varies greatly within the information sciences (Agarwal, 2005a). The main distinction lies in the use of ontology as a philosophical discipline on the one hand, and ontology in information systems engineering on the other hand.

2.2.5.1 Types of ontology

Ontology understood as a philosophical discipline deals with the nature and organisation of reality (Guarino & Giaretta, 1995). It tries to explain reality by breaking it down into concepts, relations and rules (Agarwal, 2005a). Classically, ontology assumes a *realist view* and is seen as independent of epistemology, i.e., the tenet is that since there is only one reality, there can only be a single ontology (Smith, 1998).

Ontology in information systems engineering is seen as an engineering artefact and is commonly defined as an explicit specification of a conceptualisation (Gruber, 1995). It consists of a vocabulary and a set of assumptions relating to the intended meaning of the vocabulary (Guarino, 1998). This *"partial semantic account of the intended conceptualization"* (Guarino & Giaretta, 1995, p. 26) is termed *ontological commitment*. Hence, ontology in this sense defines what can be represented in an information system. Uschold and Gruninger (1996) anticipate three benefits of taking an ontology-driven stance in information systems engineering: Improved communication between people and organisations, improved interoperability between systems, and better reliability and reusability of the developed components.

There are several typologies of ontologies. Uschold and Gruninger (1996) propose a classification according to the degree of formalisation into:

- Highly informal ontologies: Expressed in loosely natural language.
- Semi-informal ontologies: Expressed in a restricted form of natural language.
- Semi-formal ontologies: Expressed in an artificially formally defined language.
- Rigorously formal ontologies: Meticulously defined terms with formal semantics, theorems and proofs.

Another distinction is made by Guarino (1998) according to the degree of generality into the levels listed below, while each level builds on concepts defined on the higher level(s) (Figure 2.7):

- Top-level ontologies: Describe very general concepts like space, time, and event, which are independent of a particular problem or domain.
- Domain ontologies and task ontologies: Describe the vocabulary related to a generic domain (like medicine, or automobiles) or a generic task or activity (like diagnosing or selling).
- Application ontologies: Are ontologies engineered for a specific use or application focus, such as diagnosing cancer. Guarino (1998) suggests building application ontologies by integrating and specialising domain and task ontologies.

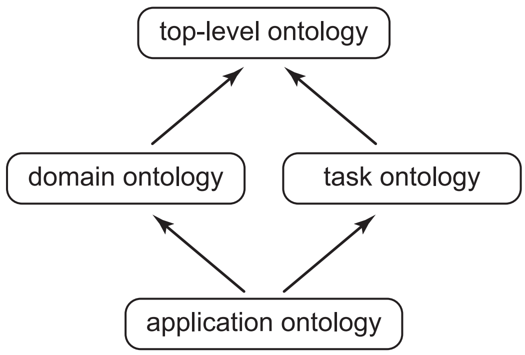


Figure 2.7: Tiers of ontology (Guarino, 1998, p. 9)

2.2.5.2 Grounding ontologies

Ontological commitments have to be made explicit, i.e., the links between the basic concepts in an ontology and the real world have to be defined. This process is termed *grounding* (Scheider et al., 2009). Several grounding methods were applied in geographical information science. Kuhn (2001) proposes a method for geographic ontologies that uses text analysis to elicit concepts in a domain from an activity-oriented perspective. Bennett et al. (2008) propose to ground ontologies in actual data. Their approach builds on rigorous formal definition of geographical concepts which can be used to extract corresponding entities from data in an ad hoc manner. Finally, Kuhn (2004) proposes to base groundings on *cognitive semantics*. While the realist view assumes that meaning 'is out there', cognitive semantics claims that meaning is incorporated in mental structures (*cognitive models*) that are shaped through perception (Gärdenfors, 1996). Gärdenfors defines *conceptual spaces* as a framework for representation of cognitive semantics. A conceptual space consists of a number of *quality dimensions*, such as weight, temperature, and area. According to Gärdenfors, admissible realisations of a concept correspond to convex regions in a concept space. Gärdenfors also explicitly makes a link to prototype theory by stating that prototypes are central points in concept space. Raubal (2005) demonstrates the utility of conceptual spaces for measuring similarity of concepts and achieving interoperability.

2.3 Analysis of urban places

2.3.1 What is an urban place?

There are various ways of defining urban places. Commonly, the following aspects are employed (Carter 1995, p. 12; Pacione, 2005, p. 22):

- Minimum population or population density
- Physical urban form, such as a contiguity of urban land use
- Presence of urban functions
- Administrative designation
- Economic criteria, for example the distribution of labour

In England and Wales, for example, an urban settlement is defined as an area having a population of more than 10,000 people (Pointer, 2005). In Switzerland urban areas are defined as individual communes having at least 10,000 inhabitants or agglomerated communes having together at least 20,000 inhabitants, whereas various physical and economic criteria are employed to establish agglomerations (Schuler et al., 2005, pp. 148–149).

While the shift from rural to urban population is still ongoing, the bulk of the population of the Western world lives in urban areas (Haggett, 2001). This raises concerns about effective design of urban space for warranting urban livelihood and limiting urban sprawl. Analysis of the configuration of urban space, and the actors and forces that drive its dynamics, contributes towards finding viable solutions.

The settlement as a unit feature of the earth's surface has two aspects: Location or position, and form or internal structure (Carter, 1995, p. 5). This thesis (and hence this review) focuses on the analysis of a city's internal structure. The urban design compendium defines urban structure as follows: *"The term urban structure refers to the pattern or arrangement of development blocks, streets, buildings, open space and landscape which make up urban areas. It is the interrelationship between all these elements, rather than their particular characteristics that bond together to make a place."* ("Urban design compendium", 2011, p. 33). Although it is acknowledged that each city is unique in its structure, cities share a number of characteristics and develop in similar ways. Conzen (1960, 1969, p. 3) analyses the *townscape*, the urban landscape, along three dimensions:

1. Land use, which marks the function of urban space.
2. The town plan, which incorporates the layout of streets and plots or urban blocks.
3. The building fabric, which relates to the architectural style of buildings.

2.3.2 Analysing urban land use

In the first half of the 20th century, a series of *ecological models* to urban land use were developed and attracted wide interest (Carter, 1995, pp. 126–139; Pacione, 2005, pp. 140–150). The ecological approach puts forward a competition for space amongst different users that eventually leads to segregation of land uses and social classes. Burgess' model of urban land use divides space into four concentric rings around the central business district (Figure 2.8a).

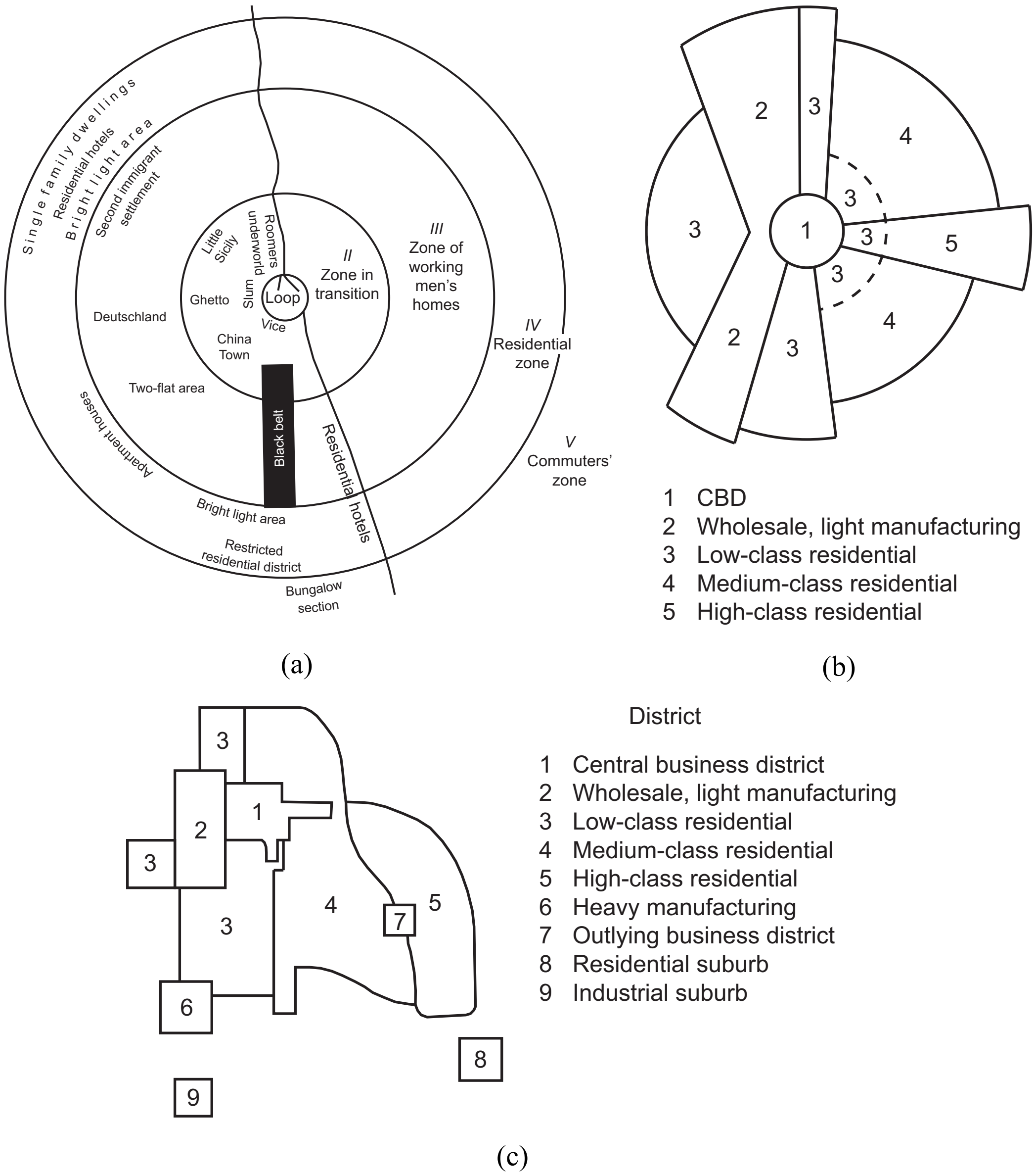


Figure 2.8: General schemes of urban land use (Pacione, 2005, pp. 242–245) (a) Burgess' concentric-zone model (b) Hoyt's sector model (c) Harris and Ullman's multiple-nuclei model

Burgess' model was modified by Hoyt who focused mostly on distribution of housing and observed that the patterns rather arrange in sectors than in concentric rings (Figure 2.8b). This model accounts for spatial inequalities such as communication routes, along which commerce and industry develops, and variations in landscape qualities such as hills, where higher class residences develop. The multiple-nuclei model by Harris and Ullman suggests that cities do not grow around a single core, but are formed by integration of separate nuclei (Figure 2.8c). Mann (1965) combines the models of Burgess and Hoyt in his model of a typical medium-size British city (Figure 2.9). It also incorporates climatic elements by assuming prevalent wind from the west, which causes inferior living conditions in the east due to industrial exhausts. The analysis of social segregations in 19th century Liverpool by Lawton and Pooley (1976) can be seen as a realisation of the sectoral model (Figure 2.10).

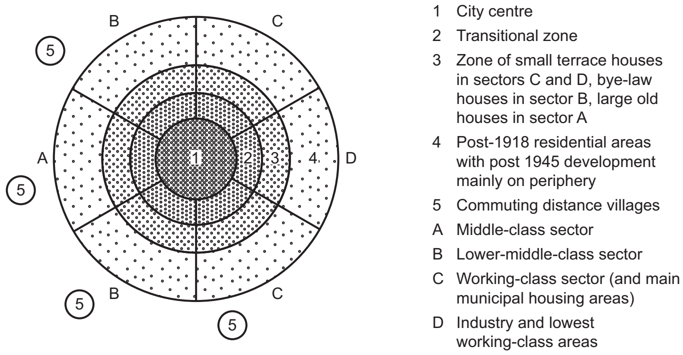


Figure 2.9: Mann's model of a typical medium-size British city (Mann, 1965, cited in Pacione, 2005, p. 247)

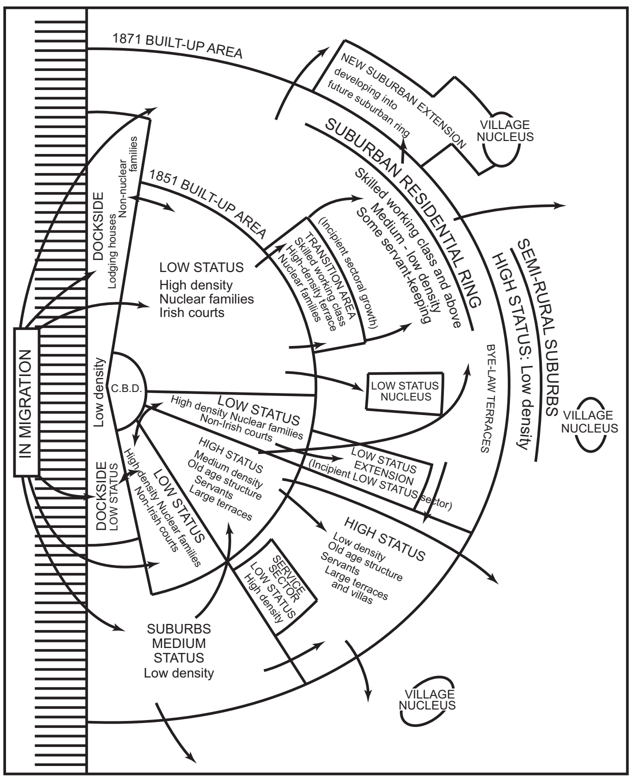


Figure 2.10: The structure of Liverpool in 1871 (Lawton & Pooley, 1976, cited in Pacione, 2005, p. 55)

The ecological models were criticised for their mechanistic view and economic bias (Carter, 1995, p. 136). The morphology of a city, as a dynamic phenomenon, forms through innumerable decisions of many different actors on the urban stage—governments, urban planners, companies, inhabitants, and many more. The influence of such individual decisions on urban evolution is systematically explored through simulation. Two modelling paradigms are dominant: Cellular automata (CA), and multi-agent systems (MAS) (Batty, 2005; Benenson & Torrens, 2005). CA model space as sets of spatially stationary cells. Each cell is an individual automaton that exhibits some properties, such as land use. Time is emulated as a series of discrete time steps, and dynamic behaviour is achieved through transition rules that determine alterations of cell properties at each time step. MAS abolish the restriction of spatial stationarity, although there can be spatially fixed agents as well (for example a parcel of land). CA and MAS are commonly employed to simulate urban land use change and urban sprawl, although there are limited possibilities for evaluation (White & Engelen, 2000).

2.3.3 Town plan analysis

The town plan is the physical manifestation of urban processes. The seminal work in town plan analysis in Britain was M. R. G. Conzen's study of Alnwick (Conzen, 1960, 1969). Conzen established town plan analysis as an integrated study of the elements that make up a town plan—street layout, plot layout, and building footprints—in the course of history. Hence, most studies in this field form detailed narratives of the historical development of an individual site and are thus hardly generalisable. However, by studying individual actors that influence formation of a townscape and their motivation (Whitehand & Whitehand, 1984) an important contribution to town planning is provided (Whitehand, 1992).

Space syntax is a research field that aims at descriptions of configured, inhabited spaces in such a way that their underlying social logic can be enunciated (Bafna, 2003). Space syntax can be used to analyse space at all scales, including building layouts, neighbourhoods, settlements, and regions. The basic tenet is that, since society and space influence each other in a reciprocal relationship, social organisation is reflected in the configuration of space (Hillier & Hanson, 1984, p. 26–27). Space is abstracted by focusing on its topology. A common technique to do this is to discretise it into a number of convex spaces (Hillier & Hanson, 1984, p. 91), and then draw a map of longest straight lines that pass through the convex spaces, called axial map (Figure 2.11). For both convex map and axial map a number of descriptive measures are proposed. Several studies demonstrate that spatial configuration as quantified by space syntax shows a striking correlation to pedestrian and vehicular movement patterns (Hillier et al., 1993; Penn et al., 1998). A possible explanation is given by Penn (2003), who points out a link between space syntax and spatial cognition. It was also proposed to combine town plan analysis in the Conzenian tradition and space syntax for achieving a more comprehensive analysis (Griffiths et al., 2010).

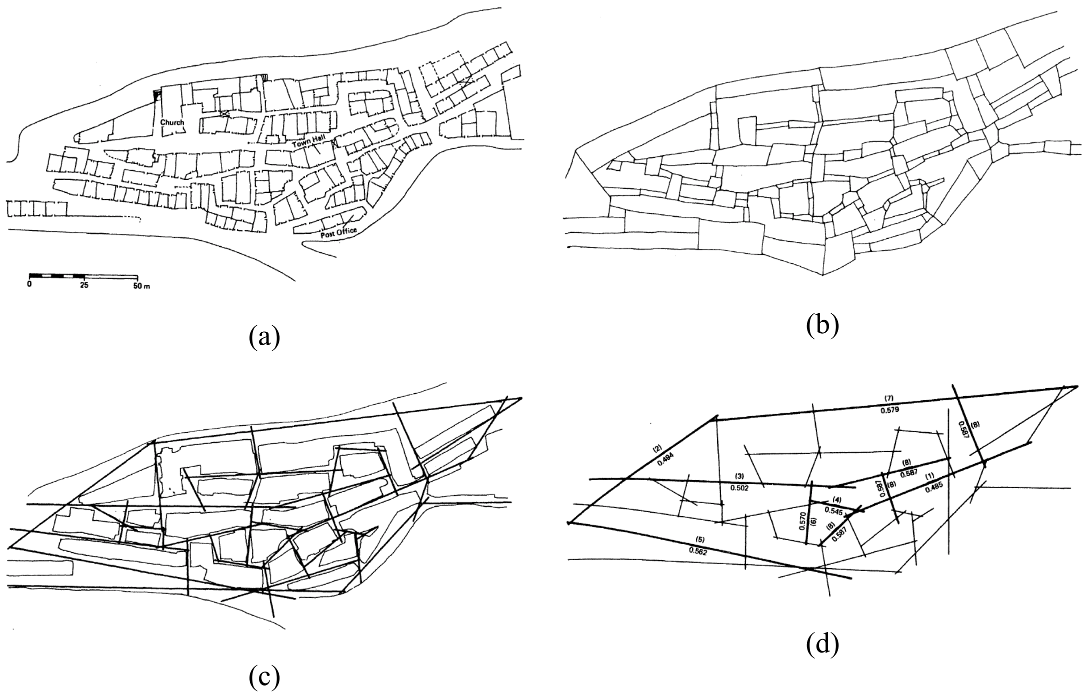


Figure 2.11: Analysis of a town layout by means of space syntax: (a) Original town plan (b) Convex map (c) Axial map (d) Axial map with the 25% most integrating (i.e., most accessible) spaces (Hillier and Hanson, 1984, pp. 90–115)

A seminal work about spatial cognition of urban environments and its relation to human wayfinding is Kevin Lynch's *The Image of the City* (Lynch, 1960). Lynch's model differentiates between five general classes of urban structural elements (Lynch, 1960, pp. 47–48):

1. *Paths* are the channels along which an observer can move. They may be streets, walkways, transit lines, canals, railroads.
2. *Edges* are linear elements that form boundaries: Shores, railroad cuts, edges of developments, walls.
3. *Districts* are medium-to-large sections of the city, which are recognizable as having some common, identifying character.
4. *Nodes* are foci to and from which an observer is travelling, such as junctions, places of a break in transportation, or a crossing of paths.
5. *Landmarks* are easily identifiable objects which serve as external reference points. A landmark can be a building, monument, sign, store, etc., which has a distinct characteristics.

The empirical basis of this model is provided by surveys of human perception of Boston, Jersey City, and Los Angeles, involving techniques such as drawing sketch maps, and describing different parts of the city, and field analysis by instructed people. However, as can be seen in the list above, the town plan plays a central role in Lynch's model.

2.3.4 Urban space and place

Place is a primary element in human structuring of space. A room, home, a park, a neighbourhood, a city, a national state all are instances of place. Although place is a common-sense notion, it is reported to be a contested concept and hard to define (Cresswell, 2004; Bennett & Agarwal, 2009). However, most writings on place focus on meaning and experience (Cresswell, 2004), conceiving place as space infused with human meaning (Couclelis, 1992), or as centres of meaning to individuals or groups, created through experience (Tuan, 1975). Beyond mere physical and functional structure, place hence encompasses aspects of feelings, activities and history. Agarwal (2004) investigated the link of place to neighbouring spatial concepts and was able to show that location, district, and neighbourhood are all kinds of places, whereas place itself is a subtype of region. One of the most important characteristics of place is its role as means of containment: Places afford a feeling of 'being inside', and other objects are located with reference to places (Bennett & Agarwal, 2007).

2.4 State of the Art: Characterisation of urban space in cartography

While the previous section discussed analysis of urban structures in a broad context, this section focuses on particular techniques that were developed in a cartographic context, i.e., based on topographic (vector) data, and on an urban context only. Many of these techniques were specifically developed for automated map generalisation (cf. Section 2.1).

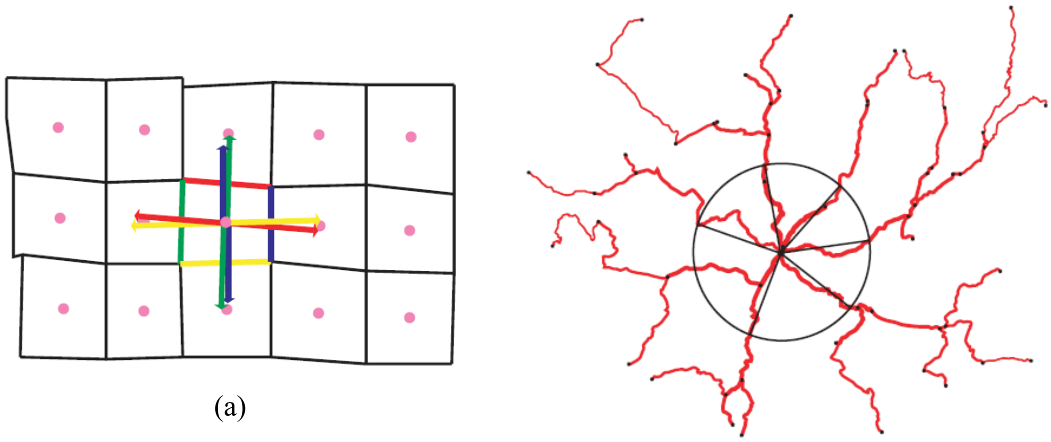
The following review of urban pattern recognition approaches is divided into approaches for characterising urban road networks, arrangements of buildings, characterising urban neighbourhoods, and modelling settlement extents.

2.4.1 Characterising road networks

Anders (2007) describes a set of algorithms for detecting different types of urban road patterns (summarised in Heinzle & Anders, 2007), aiming mainly at typification of road networks for automated generalisation. We summarise here her algorithms for detecting (rectangular) grid structures, star structures, and ring roads.

Anders' algorithm for detecting grid structures uses road meshes, which are areas enclosed by roads (inside of urban areas they are also referred to as urban blocks). It basically works by shifting centroids of candidate meshes along the edges (Figure 2.12a). If certain criteria are met (i.e. the centroid is sufficiently close to the centroid of an adjacent mesh, areas of both meshes are homogeneous, and the merged area is approximately convex), it is considered to be a grid cell.

The algorithm for detecting ring structures calculates for each node the shortest path to all other nodes in the road network. The shortest paths are then intersected with a circle around the node (Figure 2.12b). If the length of the shortest path is sufficiently close to the radius, it is added to a list of rays. If there are at least five rays that are well distributed, a star structure was found.



(a) (b)

Figure 2.12: Approaches to detect grid and ring structures in road networks. (a) Shifting of road mesh centroids to detect grid cells (Anders, 2007, p. 58) (b) Intersection of shortest paths with circle for detecting rays (Anders, 2007, p. 68)

Extraction of ring roads is based on road meshes again (Figure 2.13). Meshes are merged in a combinatorial way. For each combination of meshes, the similarity to a circle is evaluated based on a number of similarity measures, yielding an ordered list of possible ring candidates. To reduce computational complexity, road meshes are first aggregated to larger units.

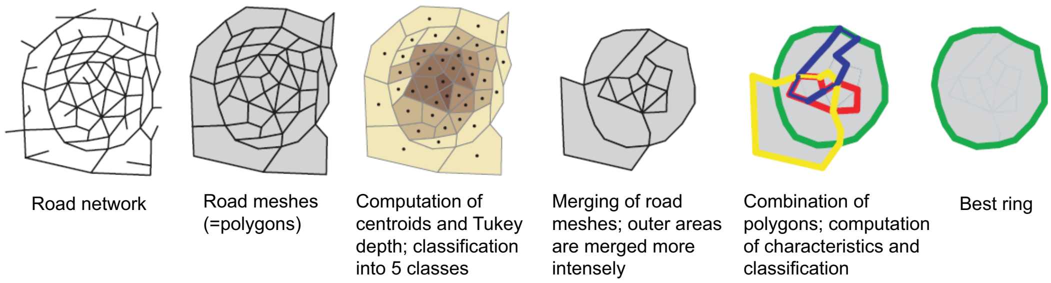


Figure 2.13: Computation of ring roads (Anders, 2007, p. 81)

Alternatively, Yang et al. (2010) present a multi-criteria decision approach for detecting grid patterns in road networks. The multi-criteria decision integrates measures of consistent direction and shape similarity between adjacent road meshes, and as similarity of meshes to rectangles.

2.4.2 Characterising arrangements of buildings

To maintain the character of an urban area while generalising it, it is important to preserve the local arrangement of buildings. Thus, there is a wealth of methods for detection of characteristic groups of buildings.

Alignments are groups of buildings that are arranged in a straight line. The method to detect alignments by Boffet (2001) and Boffet and Rocca Serra (2001) first creates triplets of linearly arranged buildings, and then iteratively merges the triplets to larger groups of aligned buildings. The method presented by Christophe and Ruas (2002) projects building centroids onto a line. Clusters of close projected points are stored as possible alignments. The direction of the line iteratively changes its direction until a full circle is covered. The list of possible aligned groups is finally filtered and merged.

Regnauld (1996, 2001) presents a graph-based method to create perceptual groups of buildings. First, a minimal spanning tree (MST) is generated containing all buildings. The MST is then iteratively segmented by eliminating edges which make the subgroups most homogeneous. A related approach is introduced by Anders et al. (1999). A relative neighbourhood graph (RNG), which is a sub-graph of the Delaunay triangulation, is computed from building centroids. A clustering algorithm is employed to remove some of the edges. The mean distance of a node to all adjacent nodes in the Delaunay triangulation is used as similarity measure for the clustering. Anders et al. also argue that using different thresholds for the similarity measure, structures of different sizes can be detected, e.g. building groups, neighbourhoods, settlements, and regions.

2.4.3 Characterising urban neighbourhoods

Barnsley and Barr (1997) and Barr et al. (2004) examine the separability of urban land use classes using graph-based structures. Land use is, unlike land cover, an abstract concept that involves aspects of form and function. Their approach requires a land cover map, which can be generated automatically from high resolution (1–5m) remotely sensed imagery. In the latter work, they manually delineated homogeneous urban neighbourhoods, which were mostly residential developments of different periods of construction. The analysis uses and classifies individual buildings. By employing the measures area, compactness, Gabriel graph edge length and node degree, they are able to show that many of the defined land-use classes are well separable, while the distinction 1950s vs. 1960s, and 1960s vs. 1970s settlement is problematic.

Steiniger et al. (2008) perform a classification of urban neighbourhoods into 'Inner City', 'Urban', 'Suburban', 'Industry/Commercial', and 'Rural' areas. They aim primarily at topographic mapping. Firstly, several topographic maps use areal tinting to reveal the structure of urban areas. Secondly, such a classification can be used to parameterise algorithms for automated generalisation (Steiniger et al., 2010). As in Barr et al., individual buildings are classified. However, a total of nine morphological measures are employed, and instead of a graph structure, buffers around each building yield context information. Finally, a supervised classification is carried out, whereas the authors compare the effectiveness of several algorithms, such as Support Vector Machines.

Another cartographic approach for characterising districts is presented by Boffet and Coquerel (2000) and in more detail in Boffet (2001). They start by creating and characterising urban blocks, which are areas bounded by roads. Then, buildings inside each block are statistically analysed regarding function, average building size, and building density. The classification into distinct groups (Figure 2.14) is done by applying predefined thresholds.

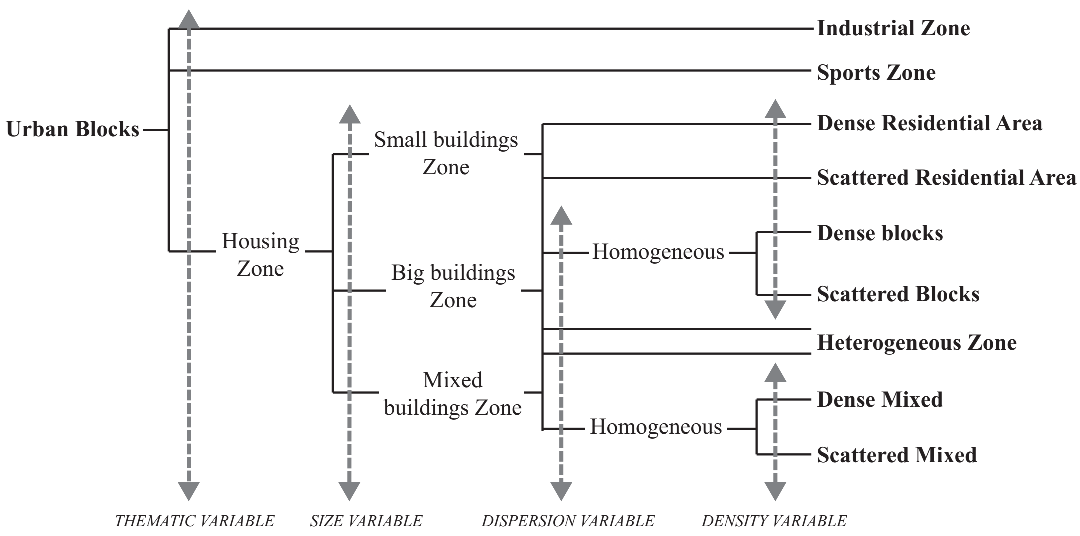


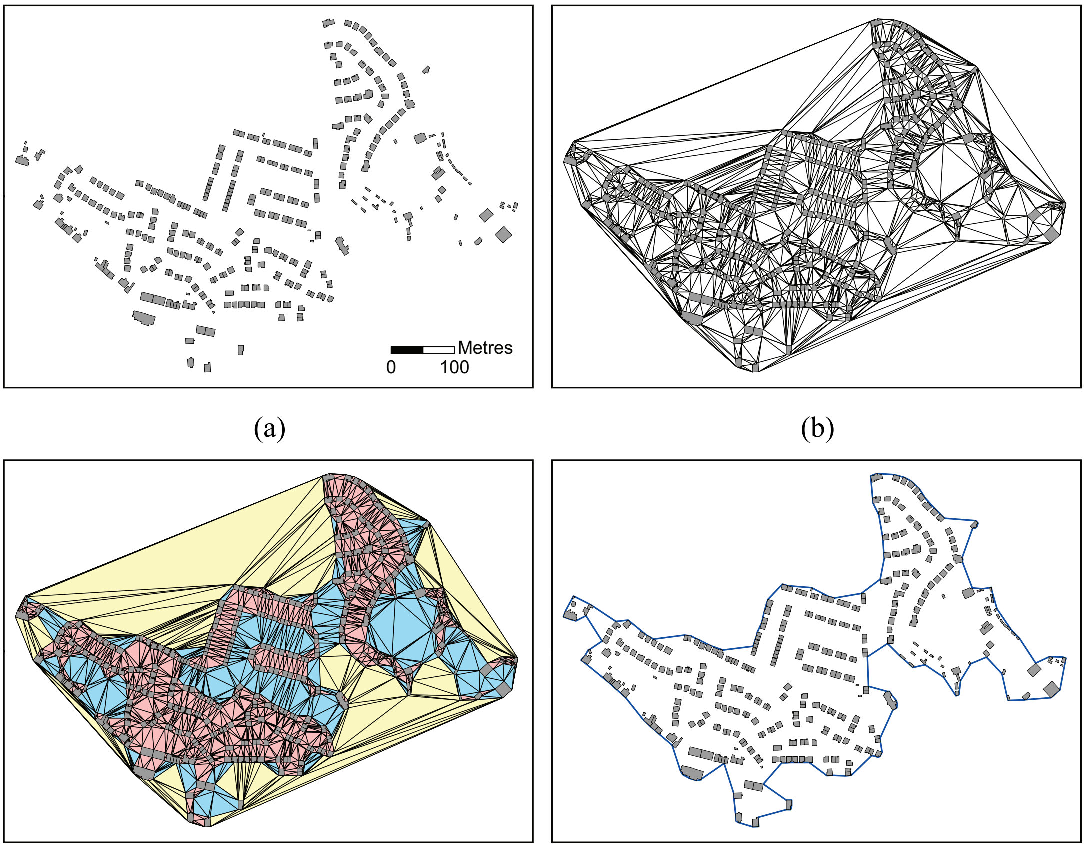
Figure 2.14: Taxonomy of urban blocks for generalization by Boffet and Coquerel (2000)

The urban block types that were found in this way are further aggregated to districts. Boffet (2001) proposes two methods. The first method simply merges adjacent blocks of similar classification. The second method uses homogeneous blocks as seed points for an iterative growing procedure, which adds at each step the most similar adjacent block to each nucleus until all blocks are assigned to a district. Similarity is measured in terms of average building size and building density.

Boffet (2001) also observes that the building density is highest in the city centre, at least for those cities having a historic core (Boffet, 2001, p. 168). Thus, she proposes to define a threshold on the building density for urban blocks, or districts, to determine the city centre.

2.4.4 Modelling settlement extents

Joubran and Gabay (2000) propose a graph-based method for modelling settlement extents, departing from a Delaunay triangulation of building ground plans, building centroids, or roads. Considering the distribution of edge lengths, a threshold for the edge length in the Delaunay triangulation is set. All edges above that threshold are removed. For remaining edges, a circumscribing hull is created. The approach is illustrated in Figure 2.15.



(c) (d)

Figure 2.15: Generation of settlement extents after Joubran and Gabay (2000). (a) Initial buildings (b) Constrained Delaunay triangulation of building ground plans (c) Elimination of edges using different thresholds (d) Circumscribing hull after elimination of edges. OS MasterMap data Ordnance Survey ©Crown Copyright. All rights reserved.

Boffet (2001) presents a density-based approach to delineate settlements, which is again working on the arrangement of buildings. In the first step, buildings are enlarged using a buffer operation. The optimal buffer size was determined through experimentation and set to 25 m. Then, buffers are merged to a settlement area. The procedure is repeated once by enlarging settlement areas obtained in the last step. This is to merge areas that are separated e.g. by highways. Finally, the outlines of the resulting shapes are simplified by dilation and erosion of the settlement areas (cf. Figure 2.16), and by applying the Douglas-Peucker algorithm (Douglas & Peucker, 1973). Based on the area of the obtained settlements they are classified into villages, small and large cities.

A very similar approach is used by Regnauld and Revell (2007) to detect urban areas and rural building clusters. Commenting that the approach generates some unwanted spikes at the boundary of settlements, Chaudhry and Mackaness (2006, 2008) refine the approach introduced by Boffet (2001). A gravity-based formula is used to model the local density of buildings and subsequently determine the buffer size for expanding the buildings.

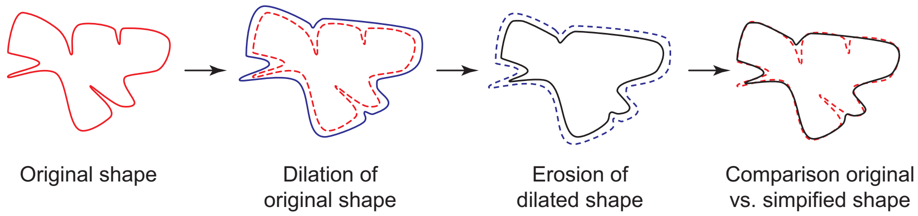


Figure 2.16: Simplification of polygon outline by dilation and erosion (Boffet, 2001)

Anders (2007) introduces a density-based method to delineate settlement extents that uses road meshes. The basic observation is that road meshes have a smaller area within settlements. Hence, a threshold is defined to extract street meshes in urban areas.

2.4.5 Semantic modelling techniques for generalisation of spatial data

With the methods presented so far, knowledge about geographical phenomena is embedded within the algorithmic recognition procedure. However, with increasing prevalence of spatial information the need to better model the semantics of represented phenomena came up as an important issue. This has promoted approaches where semantics is modelled separately of the recognition procedure, either formulated in languages that allow recognition through standardised reasoning processes, or to be converted into algorithmic representations afterwards.

One of the first representatives of this class of approach was presented by Sester (2000). The principle is to represent semantics as a network of geographic feature classes, whereas each class is characterised through some spatial properties (e.g. area, elongation) and relations (e.g. contains, parallel). A basic set of properties and relations to choose from is then provided by the system. The semantics is learned by supervised classification rather than being explicitly prescribed by a domain expert.

Greenwood and Mackaness (2002) take a partonomic view on spatial data enrichment. Their approach is extended by Chaudhry et al. (2009), who define a *functional site* as a compound entity where the relationships to its parts are made explicit. A school ground, for instance, is a functional site consisting of class rooms, playgrounds, sports facilities, etc. A method is presented that builds upon explicit modelling of partonomic relations to assemble functional sites from richly attributed topographic vector data.

Interoperability within distributed and heterogeneous environments benefits if the feature classes of spatial datasets are linked to concepts of an ontology. In this context, Klien (2007) and Klien and Lutz (2005) discuss the automatic discovery of such links through spatial data enrichment techniques. It requires that the concepts are richly described. Descriptions of concepts, which are represented in a language such as the Web Ontology Language (OWL), are converted into a series of spatial analysis functions that create extensional representations of the concepts (Figure 2.17). The extensional representations can then be used for a certain user-specific analysis, or be overlaid with feature classes to establish a similarity measure.

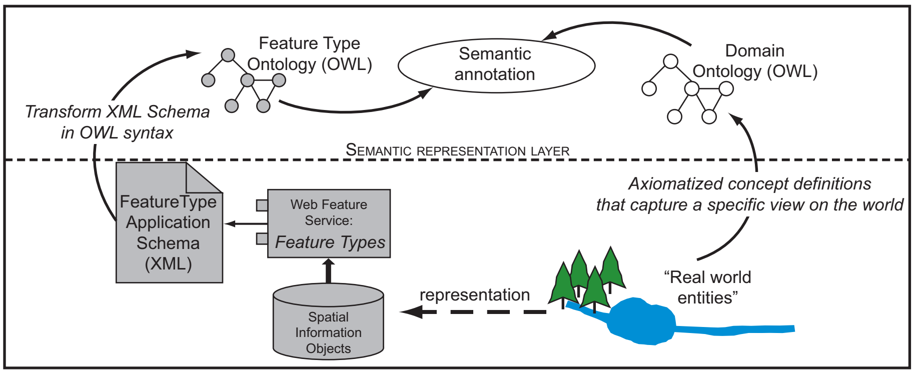


Figure 2.17: Framework for the semantic annotation of geodata (Klien, 2007, p. 440)

Mallenby (2008) presents an approach to cartographic pattern recognition with the aim to deliver user-specific representations at query-time. The approach is an extension of ideas presented by Santos et al. (2005). It was also proposed to use it to ground ontologies in data, that is providing interpretations of concepts by concrete data objects (Third et al., 2007). A three-layered architecture is proposed to handle issues of ambiguity, vagueness, and grounding in various datasets (Figure 2.18). The general layer contains high-level and context-independent definitions of concepts, such as basic spatial predicates, and commonly understood meanings of geographic concepts such as "rivers". The data layer consists of particular datasets and denotations of "basic" predicates, such as land, water, or linear. The grounding layer relates the high level concepts of the general layer to the basic predicates of the data layer.

The pattern recognition process is carried out in Prolog. The grounding layer serves as a query language to extract relevant objects for specific high level concepts. The approach takes as supervaluationist stance for handling uncertainty: Vague concepts are modelled by means of context-dependent parameters.

<table>
<tr><td colspan="3">General:
spatial and temporal logic
global structure
high-level predicates river[l](x), stream[s, l](x)</td></tr>
<tr><td colspan="2">Grounding 1:
river[l](x) $\leftrightarrow$ linear[l](x) $\wedge$ water(x)
$\neg$small[s](x)
stream[s, l](x) $\leftrightarrow$ linear[l](x) $\wedge$ water(x)
small[s](x)</td><td>Grounding 2:
river[l](x) $\leftrightarrow$ 2-linear[l](x) $\wedge$
$\exists$y(3d-bed(x,y) $\wedge$
3d-channel(y))</td></tr>
<tr><td>Data 1:
2-d topographic data

Human-scale
small[s](x)</td><td>Data 2:
2-d topographic data

Boat-scale
small[s](x)</td><td rowspan="2">Data 3:
3-d topographic /
bathymetric data

2d-linear[l](x)
3d-bed(x,y)
3d-channel(y)</td></tr>
<tr><td colspan="2">linear[l](x)</td></tr>
</table>

Figure 2.18: Layered structure relating the same general layer to multiple grounding and data layers (Third et al., 2007, p. 43)

Thomson (2009) elaborates on a methodology for capturing and explicitly representing human reasoning about topographic maps, with the aim of automatically inferring land use from the Ordnance Survey (OS) MasterMap® Topographic layer (Thomson, 2009; Thomson & Béra, 2008; Thomson & Béra, 2007). Thomson proposes to employ an ontology-driven approach to spatial data enrichment (Figure 2.19). Similar to the method presented by van Smaalen (1996), the proposed approach incrementally aggregates concepts into more abstract entities. For instance, a residential block is made up of residential houses and gardens; a couple of residential blocks make up a residential neighbourhood; a city is composed of a set of neighbourhoods (of residential and non-residential function).

The implementation of the approach builds on description logics (DL) as modelling and reasoning language. The feasibility of the approach is shown at the example of classifying types of residential housing—detached, semi-detached, and terraced houses—and aggregating them into residential districts. However, due to limitations of current DL reasoners the process required manual interaction at each stage to assert intermediate results as facts, compute properties on them, and trigger the next classification step. Thomson also observes that DL reasoning currently is not practical for reasoning with a large set of instances, does not deal with uncertainty.

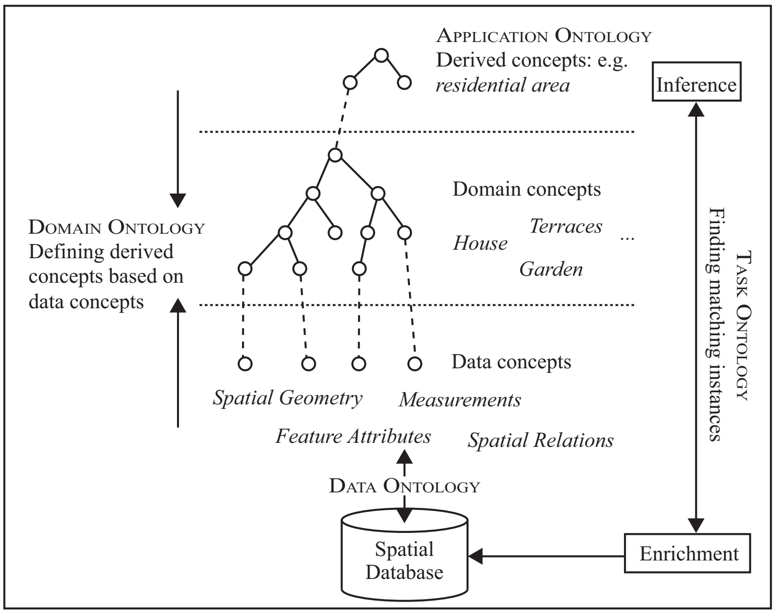


Figure 2.19: Ontology-driven approach to spatial data enrichment (Thomson, 2009, p. 145)

2.4.6 Spatial Data Infrastructures

Most of the previously discussed techniques for cartographic pattern recognition are based on geometric information only, as this was generally the only type of information that was widely available until very recently. However, there are endeavours under way in many countries to provide an integrated and harmonised access to information maintained by public administrations. The major goals of such projects are firstly to avoid unnecessary duplications in data maintenance and promote data reuse, and secondly to facilitate integrated analysis (Ellenkamp & Maessen, 2009; Murray et al., 2005). As much of public administration data is spatially oriented, Spatia Data Infrastructures (SDIs, De Man, 2006) are currently being established. The Swiss ordinance on a national spatial data infrastructure presently defines 178 datasets, encompassing topography, addresses, cadastral information, building and business registries, and many more[1]. In the Netherlands, a system of *key registries* was defined, the core of which encompasses topography, buildings, addresses, cadastral information, persons, companies, and real estate value (Ellenkamp & Maessen, 2009). A central component of the British initiative on a Digital National Framework (DNF) is the introduction of *unique identifiers* that enable linking individual features within and across datasets (Ordnance Survey, 2004). In OS MasterMap®, the TOID attribute is used to uniquely identify topographic objects. Government agencies are supposed to link their data to OS MasterMap® rather than capture their own topographic information.

[1] http://www.admin.ch. Geoinformationsverordnung, GeoIV (SR 510.620). Accessed 04.09.2011.

Consequently, there is increasingly additional information available that is linked to the topographic database and can be exploited for cartographic pattern recognition. It is thus foreseeable that in the future the challenges of semantic enrichment will change. However, it is equally foreseeable that there will always be a role for cartographic pattern recognition as no SDI will ever provide the concepts needed by the whole range of applications.

2.4.7 Relevant work in related fields

Knowledge-based approaches have been employed for image interpretation in computer vision and remote sensing (Crevier & Lepage, 1997; Blaschke, 2010). A common denominator of these approaches is that they build on image segmentation (Blaschke, 2010), whereas more recent approaches in object-based image analysis for remote sensing integrate segmentation and object modelling at multiple scales (Burnett & Blaschke, 2003). Objects are aggregated and interpreted by means of rules describing shape, spectral reflectance, and spatial context. For example, a grassy area inside an urban area is more likely to be classified as a park than as a pasture. Knowledge-based techniques are also used to support object recognition from 3D point clouds (Dehbi & Plümer, 2011). Such methods are used to build 3D city models from laser scanning data.

2.5 Summary: Challenges for research

This chapter introduced a number of topics that are relevant for data enrichment in the urban domain. Furthermore, it presented a review of the State of the Art in data enrichment research in the relevant context. Enriching semantics of topographic datasets is of increasing importance for data providers and consumers. However, many of the currently existing algorithms were built for very specific purposes, most notably for the support of cartographic generalisation operations. Hence, they aim at the preservation of spatial patterns, and the knowledge about higher order phenomena they derive is implicitly modelled in the algorithms.

First calls for more explicitly semantics-driven approaches were made almost twenty years ago (Nyerges, 1991). The fact that there are yet few examples of complete applications in this area bears out the complexity of this task. This thesis contributes to the current body of knowledge by addressing the following challenges:

1. *Focus on phenomenological approach.* Abstraction of spatial databases over large changes of scale and for specific purposes is essentially a modelling problem—it requires understanding how geographic phenomena act together when forming a higher order phenomenon. In this context, there is yet little knowledge about appropriate modelling techniques of phenomena and the composition process.
2. *Semantic grounding.* A further issue is the knowledge elicitation process for modelling higher order phenomena. This is known as the knowledge acquisition bottleneck (Weibel et al., 1995). Approaches for solving the bottleneck are known as knowledge engineering. This thesis demonstrates the usefulness of two different knowledge engineering approaches for building models of higher order phenomena: Literature analysis, and human subject experiments.
3. *Computational efficiency.* The size of spatial datasets and the complexity of pattern recognition dictates that the recognition procedure is computationally efficient. Existing proposals for ontology-driven pattern recognition stay on a conceptual level (Klien, 2007), or were not carried out on large amounts of data and relatively simple problems (Thomson, 2009).

In response to the challenges discussed above, this thesis develops a framework for ontology-driven recognition of urban structures. The framework is a top-down approach, comprehending knowledge engineering, knowledge modelling, and computationally efficient implementation. The framework is illustrated on two urban phenomena, and validated based on large datasets.

Chapter 3
Summary of Papers

The four research papers included in Part II represent the main substance of this thesis. This chapter summarises the research papers for providing a basis to the subsequent discussion in Chapter 4. For each paper, the objectives, methods and results, and contributions are highlighted. These summaries, however, do not provide a substitute for the reading of the full papers.

3.1 Paper 1: Developing an ontology-driven methodology

Lüscher, P., Burghardt, D., & Weibel, R. (2007). Ontology-driven Enrichment of Spatial Databases. *10th ICA Workshop on Generalisation and Multiple Representation*, Moscow, Russia, August 2–3, 2007.

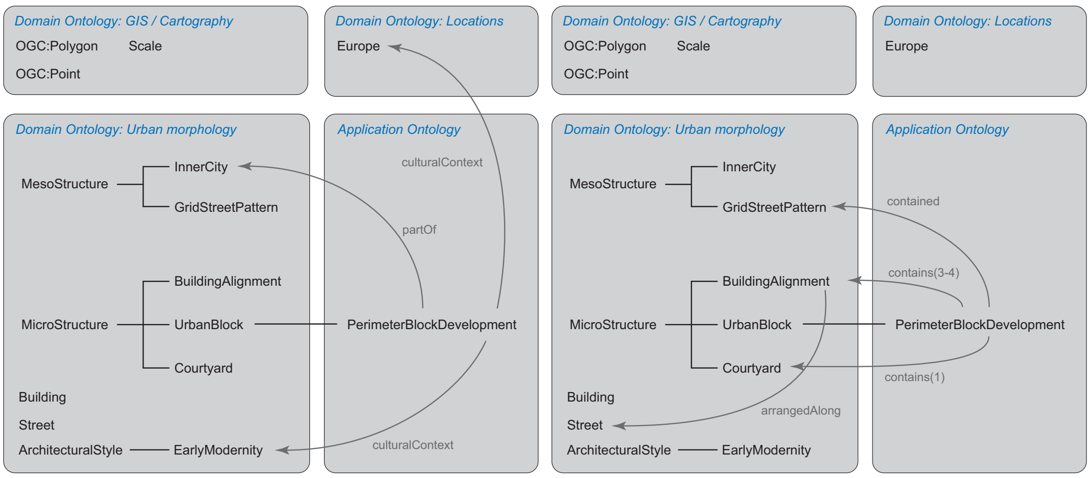
3.1.1 Objectives

The first paper motivates the development of an ontology-driven methodology to spatial data enrichment. The paper indicates shortcomings of purely algorithmic approaches, and analyses requirements and research challenges of an ontology-driven approach. It thus introduces the framework which is further developed in the subsequent papers.

3.1.2 Methods and results

Extracting very specific concepts from spatial databases and achieving drastic abstractions in map generalisation require that complex semantics is modelled explicitly in the pattern recognition process. The paper identifies shortcomings in purely algorithm-driven approaches as they are often oriented towards visual optimisation (rather than geographic meaning), hide assumptions about modelled patterns and are monolithic. To mitigate these shortcomings, the paper discusses a top-down approach that captures semantic models in ontologies and uses these models to drive the pattern recognition process. The paper subsequently presents first ideas on semantic modelling and execution of the pattern recognition process, and discusses challenges for research.

For conceptual modelling, a distinction is made between domain ontologies and application ontologies. The former is a library of generic concepts that can be used to explain specific urban concepts in the application ontology. The paper also introduces two different aspects of concept knowledge for map generalisation: Cultural context knowledge allows flexible abstraction, while additional knowledge of spatial characteristics is needed for pattern recognition (Figure 3.1).



(a) (b)

Figure 3.1: (a) Geographical and cultural context and (b) spatial characteristics of a perimeter block development

The paper draws upon Description Logics (DL) reasoning to carry out the pattern recognition process. Due to limitations of DL reasoning when dealing with fuzzyness and large quantities of data, later works translate the conceptual model directly into a pattern recognition workflow.

3.1.3 Main findings and contributions

- A number of individual algorithms for recognition of urban structures exist in generalisation research. However, they are most often developed for solving very specific problems in cartographic generalisation and detached from human experience of urban space.
- To overcome limitations of the aforementioned algorithms, a top-down framework for data enrichment is proposed. The framework provides the basis for the subsequent Papers 2–4, which study aspects of it in more detail.
- The paper states a number of challenges of ontology-driven pattern recognition that need further study. The transition between abstract models of urban structures and the pattern recognition algorithm needs further study in order to determine whether it can be made more automated. Operationalisation of data enrichment processes also requires that issues of vagueness are considered.

3.2 Paper 2: Conceptual models from expert knowledge

Lüscher, P., Weibel, R., & Mackaness, W. (2008). Where is the Terraced House? On The Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases. In A. Ruas & C. Gold (Eds.), *Headway in Spatial Data Handling. Proceedings of the 13th International Symposium on Spatial Data Handling* (pp. 449–466). Berlin / Heidelberg: Springer-Verlag.

3.2.1 Objectives

The second paper shows how concept definitions for cartographic pattern recognition can be acquired from literature describing a specific domain. Furthermore, it provides a proof-of-concept for ontology-driven pattern recognition by performing the complete workflow at the case study of English terraced houses, as they are defined in the urban morphology literature.

3.2.2 Methods and results

The paper presents a step-by-step methodology for capturing semantics of geospatial concepts and using this semantics to drive the pattern recognition process. The four individual stages of the methodology are as follows (Figure 3.2):

1. In the first stage, knowledge about urban structures is acquired. In the presented case study, literature on urban morphology is used to this end.

2. This knowledge is formalised in the second stage. An *ideal prototype* of an urban structure is defined by relating the structure to other concepts, such that its meaning can be derived from the meaning of the related concepts (which are called lower order concepts in this context). *Concept maps* are used as a graphical instrument for knowledge modelling and communication (e.g. with domain experts).
3. In the third stage, algorithms are triggered to detect instances of the prototype in a spatial database. The paper introduces a formalism to turn the concept maps into a data enrichment process. It assumes that each lower level concept can be assigned a congruence value to its prototype. The congruence value of a complex concept is then calculated by weighted summation.
4. In the final stage of the ontology-driven data enrichment framework, the enriched semantics is transferred back to the database.

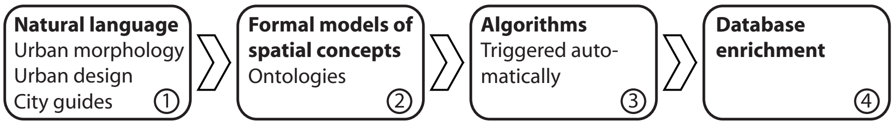


Figure 3.2: Stages in the processing chain of ontology-driven enrichment

As a proof-of-concept of the approach, the urban morphology literature is analysed for descriptions of urban residential house types (Figure 3.3). The steps of the approach are carried out to model a prototype of the English terraced house, and detect instances of terraces in OS MasterMap® data.

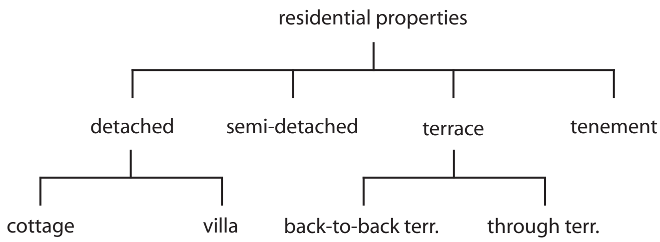


Figure 3.3: Urban residential house types extracted from the glossary of urban form

3.2.3 Main findings and contributions

- Literature analysis can be used to ground semantic models for cartographic pattern recognition.
- Vagueness of geospatial concepts is accounted for by introducing congruence values to ideal prototypes.

- Selection of algorithms that implement measures is not trivial since there are a variety of possibilities to implement concepts such as linear arrangement, or express regularity, with different outcomes.

3.3 Paper 3: Vague reasoning for concept definitions

Lüscher, P., Weibel, R., & Burghardt, D. (2009). Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. *Computers, Environment and Urban Systems, 33*(5), 363–374.

3.3.1 Objectives

Descriptions of urban structures have to be converted to a series of spatial analysis operations in order to detect them in spatial databases. Previous works proposed to use description logic reasoning (Thomson, 2009) or rule-base inference (Klien, 2007) to this end. The ambition of ontology-driven data enrichment is to encode generic spatial analysis operations in algorithms, while generic reasoning engines compose the generic algorithms to a pattern recognition workflow.

Geospatial phenomena exhibit various types of uncertainty, which is unaccounted for in previous approaches. The first objective of the paper is to show how uncertainty can be respected in pattern classification tasks by turning the concept map into a Bayesian network. Bayesian networks have previously been proposed to resolve uncertainty when reasoning about ontologies. The second objective of the paper is to explore to which extent generic spatial analysis operations can be used to detect urban structures in a realistic setting.

3.3.2 Methods and results

The concept map follows the idea of phenomenological modelling and explains the semantics of a phenomenon in terms of different types of concepts and relations. A *complex concept* is a phenomenon definable by relations (e.g., `terraced house` and `yard` in Figure 3.4). *Abstract concepts* are to be implemented algorithmically as they constitute general units that are inefficient to break up further (e.g., `row of houses`). Relations can be either *restrictions on properties* (e.g., `hasArea`), or they can be restrictions on inter-object relationships. Inter-object relationships employed are *taxonomic* (is-a), *topological* (`adja-`

`centTo`), *partonomic* (`partOf` and `contains`), and *metric* (`presenceOf`, which is implemented as a density operation).

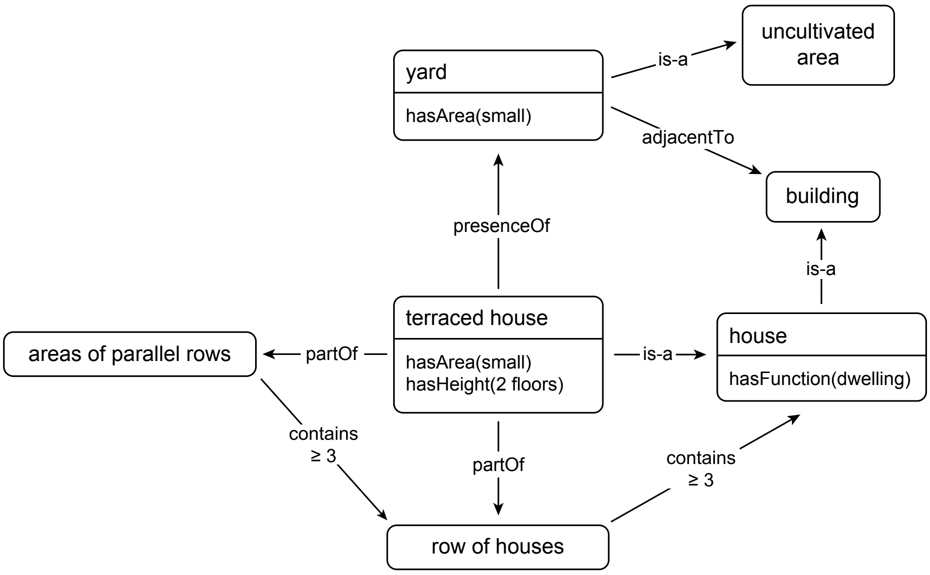


Figure 3.4: A concept map of terraced houses suited for data enrichment

For the pattern recognition process, abstract concepts and relations have to be mapped to (often spatial) operations. Many of the property restrictions are inherently vague, for instance, it is not possible to indicate a clear-cut threshold for `hasArea(small)`. Instead of assigning thresholds, the paper proposes to employ Bayesian reasoning to account for this kind of vagueness. Bayesian reasoning requires that conditional probability distributions for the decisions are known, for instance to determine he probability for an object being terraced house, given its area (and other restrictions). However, probability distributions can be estimated from training data.

To explore the second objective of the paper, an ontology of terraced houses published by the Ordnance Survey GeoSemantics team was fed into the Jena general purpose reasoning engine. It uses only crisp thresholds and topological relationships.

Both approaches were employed to classify urban areas in four English cities. It is shown that the 'simple ontology' approach produces satisfying results when only prototypical residential areas are considered, but delivers misclassifications in many other areas because the topological relationships are not powerful enough to deliver a synoptic view in all cases. Figure 3.5 shows an example of errors produced by the simple ontology approach.

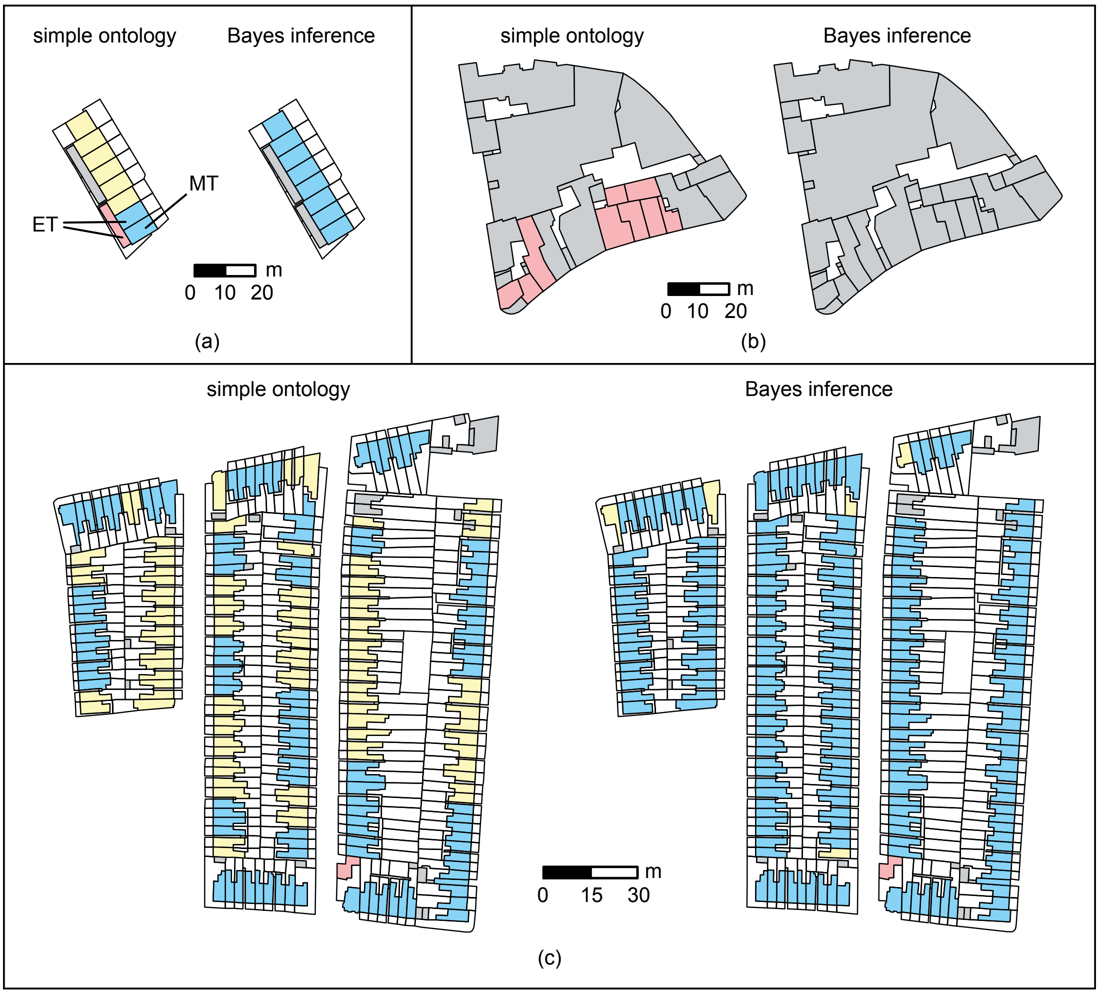


Figure 3.5: Typical errors produced by the simple ontology approach. Blue: Correct terraced houses. Grey: Correct non-terraced houses. Red: False positives. Yellow: False negatives. OS MasterMap data Ordnance Survey ©Crown Copyright. All rights reserved.

3.3.3 Main findings and contributions

- A modelling framework is presented that aligns with the ideas of phenomenological generalisation. Previous work is extended by introducing different types of inter-object relationships, in particular *metric* relationships.
- Translation from concept maps to Bayesian networks is introduced for pattern classification. Bayesian reasoning is a probabilistic approach to vagueness. It demonstrated that pattern recognition using Bayesian networks is computationally viable for large datasets.
- The paper also showed that logic reasoning techniques should best be combined with a set of algorithmic components to produce satisfying results.

3.4 Paper 4: Exploiting empirical knowledge

Lüscher, P., Weibel, R. (submitted 2010). Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases. *Computers, Environment and Urban Systems*, revised manuscript submitted June 2011.

3.4.1 Objectives

A concept such as a terraced house is 'simple' in the sense that it is defined through its physical structure. In contrast, a city centre is a concept that has a complex and vaguely defined meaning. The first aim of the paper is to demonstrate semantic grounding of such concepts by formalising empirical knowledge acquired through participant experiments. The second aim is to acquire a prototypical definition of a British city centre and develop a procedure to spatially delineate British city centres from topographic databases. Finally, the paper shows the benefit of including detailed functional information into the model generalisation process, rather than constraining on geometric aspects of space. Functional information was obtained in the form of points of interest datasets.

3.4.2 Methods and results

Figure 3.6 shows an overview of the procedure employed in the paper to delineate city centres. A participant experiment was carried out to gain a solid basis for the physical and functional characteristics that constitute a British city centre. The experiment consisted of an online survey that was distributed amongst a number of British academics and through bulletin boards on the internet. Two tasks of the survey were used to define a city centre model:

- The first task was meant to obtain an uninfluenced, individual image of a city centre. Participants had to describe important features of city centres.
- In the second task, participants were presented a list of urban features and asked to decide whether the features were typical of a city centre.

101 valid answers were collected and analysed quantitatively and qualitatively to obtain a prototypical definition and computational model of a British city centre.

The procedure to delineate city centres is raster-based. The computational model consists of a set of individual typicality surfaces that contribute positively or negatively to a perceived city centre typicality, such as 'Places to eat and drink', 'Civic services', and 'Public transport services', 'Retail parks', and 'Industrial areas'. For each class the influence on perceived city centre typicality is defined. The individual typicality surfaces were combined by weighted summation into an aggregate city centre typicality surface. Finally, a region growing algorithm was presented to spatially delimit crisp city centre areas. In the paper, city centres of ten British cities were delineated this way.

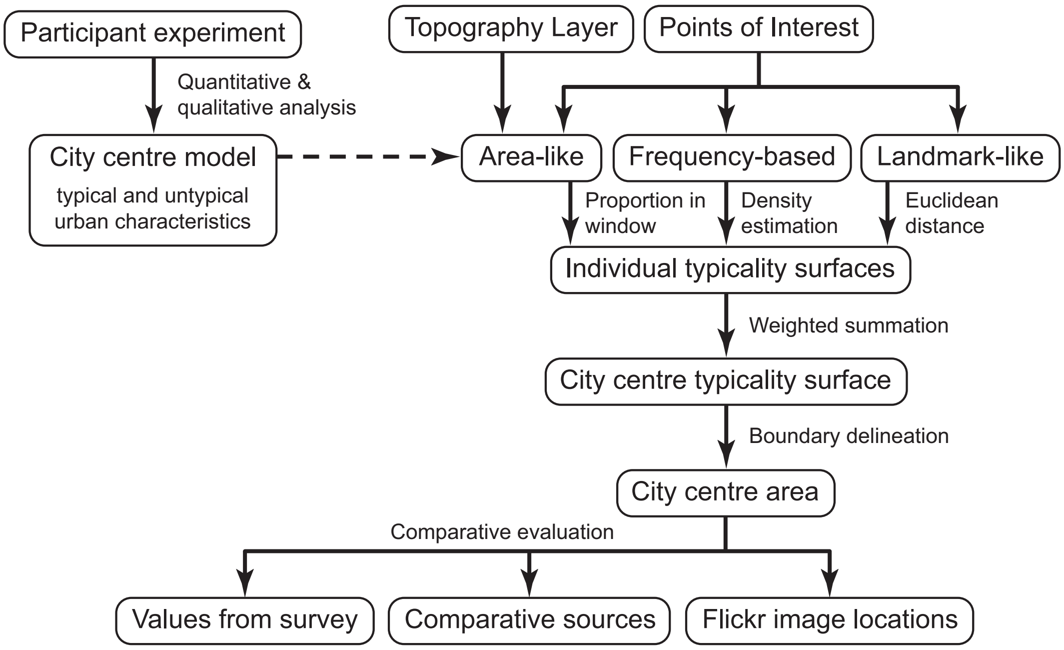


Figure 3.6: Overview of procedure for computing a city centre

Three different methods were employed to evaluate the delineated city centres. Firstly, alternative crisp representations were obtained by manually searching the internet for maps designating a city centre area and textual descriptions of city centre extents. Secondly, vague representations of city centre could be obtained for a few cities by density analysis of geo-referenced images tagged with ‘city centre’ on flickr.com (see Figure 3.7). Thirdly, a task of the participant experiment consisted of rating panoramic images for city centre typicality. Empirical city centre typicality values for 15 locations were obtained this way.

A quantitative comparison of computed city centres and alternative representations resulted in F_1-scores between 0.45 and 0.88, suggesting that the delineation produced plausible city centre areas. The computed city centre typicality values correlate well with empirical city centre typicality of panoramic image sites (r^2 = 0.916), which seems to indicate that the key functions of a city centre have been picked up by the computational model.

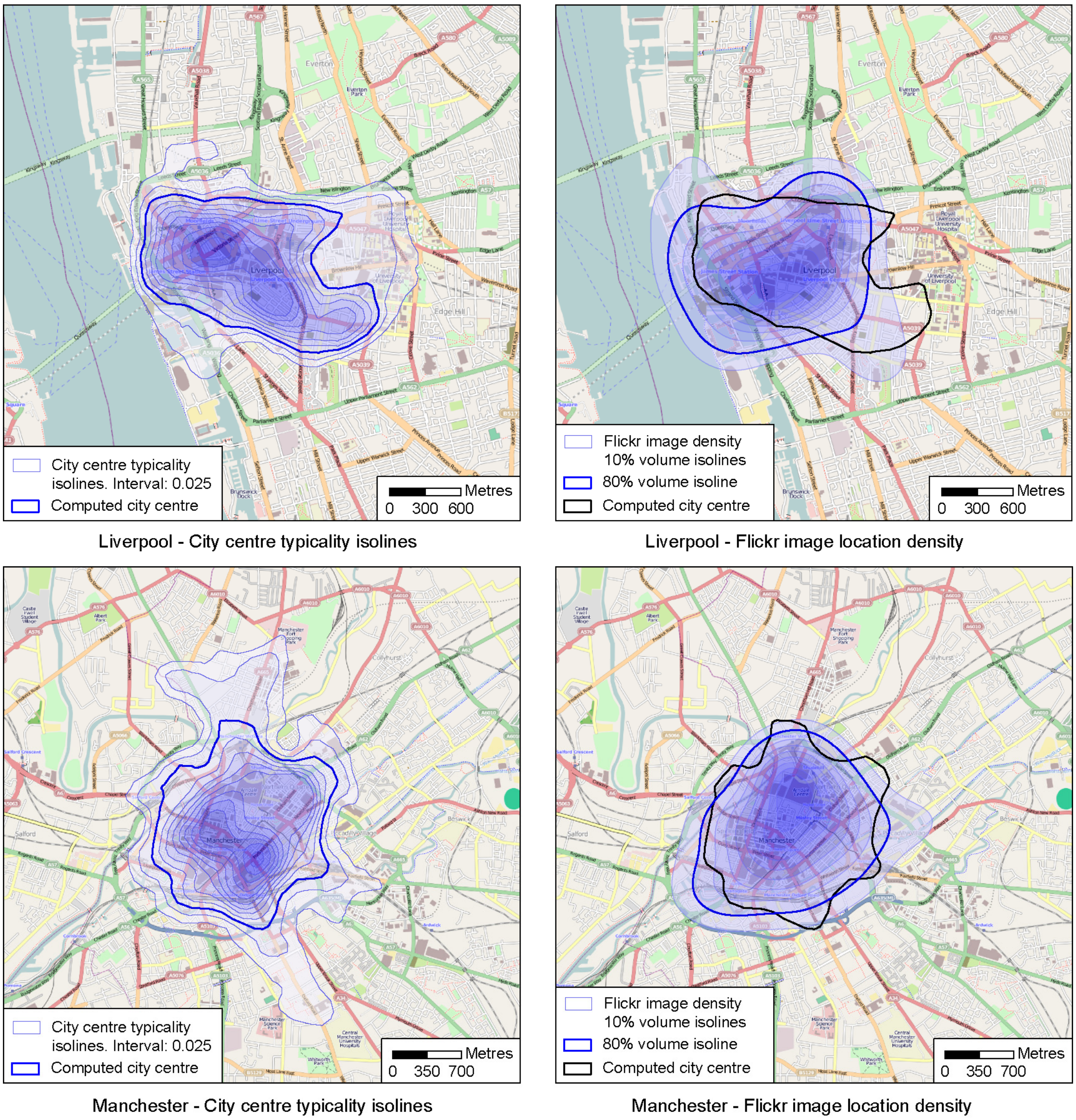


Figure 3.7: Computed city centre typicality (left) and Flickr image location densities (right) in Liverpool and Manchester. Background mapping © OpenStreetMap contributors, CC-BY-SA.

3.4.3 Main findings and contributions

- A city centre is a complex functional concept that is only vaguely defined. The main contribution of the paper is the integration of human subject experiments and pattern recognition for capturing and delineating such phenomena.
- The definitional vagueness of the concept implies that its boundaries are vague, too. The adoption of a 'city centre typicality' measure imitates the vague notion of a city centre.
- It turned out to be hard to define reference areas. Although cities define central areas for urban planning purposes, they do not match well with a vernacular representation. Ana-

lysing information from the internet offers a complementary approach. However, the main challenge is to guarantee representativeness.

- The paper showed how high level geographic concepts such as residential areas (derived through methods shown in the second paper) can be composed into yet higher order concepts.

- Carrying out data enrichment for such a complex concept also required a combination of different strategies, such matching operations for integrating POI and topography data, vector based region growing for finding industrial areas, and raster-based density algorithms. Hence, while the definition of the city centre concept is still transparent, the actual data enrichment process cannot be derived automatically from the description, and more powerful formalisms for its standardised description must be found.

Chapter 4
Discussion

There is a gap between how geography is modelled in current topographic databases and how it is used. As a survey conducted by Davies et al. (2009) revealed, professionals of various disciplines maintain that concepts related to urban area and place, such as settlement, neighbourhood, townscape, and urban structure, are key spatial concepts. Egenhofer and Mark (1995), disapproving the gap between how human users want to use GIS, and the concepts actually offered by GIS, coin the term *naïve geography* to study the body of knowledge that people have about the surrounding geographic world. They postulate that human spatial reasoning is chiefly qualitative, i.e. based on (vague) spatial regions and spatial relations.

Hence, NMAs and other data producers strive for providing higher level semantics for establishing a user-driven access to geographic information (Davies et al., 2005). A major hurdle for enriching topographic databases with higher level semantics is the lack of expertise for acquiring and modelling knowledge about geographic meaning (Nyerges, 1991).

Winter (2001) asks in his editorial whether ontologies in GIScience are a buzzword or paradigm shift, and whether they are just a modern flowery phrase for familiar concepts. He concludes that studies on ontology and epistemology are promising and go beyond conceptual modelling, since they call for grounding of models. Thus they make the commitments to the world beyond the knowledge base itself explicit.

In this research, ontology is used to denote a systematic approach to studying and describing the meaning of geographic phenomena, which is otherwise often hidden: *"Conceptualizations are often tacit, that is, they are often invisible components of our cognitive apparatus, which are not specified or thematized in any systematic way. But tools can be developed to* *render them explicit (to specify and to clarify the concepts involved and to establish their logical structure).”* (Smith & Mark, 2001, p. 593). An important role of ontologies lies in revealing the logical structure of existing conceptualisations (Teller, 2007).

This chapter provides an integrated discussion of the work presented in the four individual papers. The discussion first recalls the research questions from Section 1.2.2 that motivated this thesis. Then, the strengths and limitations of the presented methodology are evaluated in a holistic manner. Finally, the chapter concludes by listing potential applications of the research.

The discussion of research questions in the following section is structured according to the main objectives of the research set out in Section 1.2.1.

4.1 Revisiting the research questions

4.1.1 Developing a methodology to enrich spatial datasets

(I) *How can semantic modelling help in the development of cartographic pattern recognition methods?*

Paper 1 discussed the potential of an ontology-driven approach to cartographic pattern recognition for overcoming some of the shortcomings of purely algorithmic, bottom-up approaches. Pattern recognition using purely algorithmic approaches relies on specific conceptualisations of phenomena—however, often the conceptualisation is implicit in the algorithm, mixed with issues of database-specific representation, and based on observation of cartographic symbols rather than on the real-world meaning of the phenomenon. Indeed, there is little relation between the rich descriptions of urban structure discussed in Section 2.3 and the algorithms presented in Section 2.4.

Semantic modelling makes explicit the semantic assumptions behind the phenomena. The top-down methodology introduced in Paper 1 shifts the focus on modelling user-specific conceptualisations. This firstly helps to build user-driven geographic information systems. It is unrealistic that a designer of a database is able to anticipate all possible uses of the database. Hence, what is needed is a database of atomic concepts and a methodology to derive specific concepts from the database (Hart & Greenwood, 2003). Secondly, in map generalisation, achieving abstractions over large scales in detail is a knowledge-intensive process and requires that the meaning inherent among the phenomena is represented explicitly (Mackaness, 2006; Lüscher & Weibel, 2010).

Conceptual models provide better transparency to the cartographic pattern recognition process since textual or graphical descriptions can be generated upon request. The models separate the conceptual knowledge from operations needed to detect individual instances from a specific set of databases. Moreover, a description of the conceptualisation allows supporting users with automatic assessment of fitness for use by measuring concept similarity (Janowicz et al., 2010; Rodríguez & Egenhofer, 2004). For example, there is no unique definition for 'urban area'. Depending on the context of use, a definition based on function, on building density, or population density might be appropriate. However, it is important to make such decisions explicit to the user. Enhanced transparency facilitates interoperability. Hence, ontology-driven pattern recognition might also be seen as a contribution towards building Spatial Data Infrastructures (SDI, De Man, 2006).

(II) *What are the requirements for an ontology-driven approach to data enrichment in an urban context?*

Paper 1 states basic conditions for an ontology-driven approach that are needed for effective and practical application. Firstly, geospatial concepts are subject to various types of uncertainty. Hence, ontology-driven pattern recognition needs to take into account ambiguity and vagueness. Secondly, the pattern recognition procedure should be adaptable to different conceptualisations or datasets with little programming involved. Hence, the pattern recognition procedure should rely on individual building blocks that can be parameterised and rearranged. Thirdly, a comprehensive system for ontology-driven pattern recognition encompasses a user interface for browsing concept definitions and generating new concepts, including their detection in databases. And finally, practicability demands computational efficiency which means that he pattern recognition procedure should work on large datasets, and deliver satisfying results in varying situations.

4.1.2 Exploring methods for semantic grounding

(III) *What methods are available for extracting knowledge about urban structures?*

The thesis focused on modelling semantic meaningful concepts with the aim of detecting them in topographic databases. On the one hand, this requires that the concepts are to be meaningful to a human user, and on the other hand that the pattern has to be described in a level of detail appropriate for providing sufficient discriminatory power. In this context, the research applied three different knowledge extraction approaches. Firstly, a readily available formal ontology, built for facilitating interoperability, was taken as a source and tested for its potential to conduct pattern recognition (first experiment in Paper 3). The classification results suggested that the concept definition should be extended to provide enough discriminatory power for pattern recognition. Secondly, a conceptual model was extracted by analysis of a semi-formalised vocabulary about a domain (i.e., urban morphology, see Paper 2). The previous two methods stand for a class of concepts that are well defined or have a distinct meaning within a domain that can be settled on together with domain experts. Thirdly, city centre was examined as an instance of a social construct. Such concepts have a complex and vaguely defined meaning. However, there are common elements in the definition of city centre which can be acquired through human subject experiments and used for pattern recognition, as demonstrated in Paper 4. Such approaches are promising since they are a way to incorporate epistemological issues into concept definitions. In this way, they contribute to formalising multiple conceptualisations of the same reality (Schuurman, 2006).

4.1.3 Developing instruments to model knowledge and derive the data enrichment process

(IV) *Can urban structures be decomposed in terms of the phenomenological approach?*

Phenomenological modelling was shown for two case studies. The modelling was carried out in two stages. In the first stage, the concept was modelled informally by means of a concept map or concept network. In the second stage, the concept map was turned into a sequence of pattern recognition algorithms.

A first observation concerns the rather minor role of taxonomic relations in our models. This is in contrast to some traditional ways of expressing ontologies, where much of the semantics is contained in the taxonomic structure (e.g. Sen, 2007; Kavouras & Kokla, 2002). Mizen et al. (2005) discourage the use of hierarchies since many domains do not have a clear classification structure and taxonomies decrease the potential for reuse. This is evident in the case of a city centre, which has a variety of characteristics and hence cannot be placed into a classification tree unequivocally. Paper 4 describes a city centre as a subconcept of "urban district", i.e., it is an area entirely located within an urban extent (with no further details). WordNet 3.0[1] describes the taxonomic classification of city centre similarly shallowly.

[1] http://wordnet.princeton.edu/. Accessed 14.02.2011.

WordNet organises a taxonomy composed of synsets, which are sets of synonymous terms. The classification tree of city centre is shown in Figure 4.1. It can be seen that WordNet relates city centre to abstract hypernyms. It can hence be argued that relations other than taxonomic ones are more relevant for defining urban features, which creates a network-like structure (rather than a tree structure).

```
location ("a point or extent in space")
└ region ("a large indefinite location on the surface of the Earth")
  └ area ("a particular geographical region of indefinite boundary")
     ├ broadcast area
     ├ disaster area
     ├ hunting ground
     ├ neighbourhood
     ├ block, city block
     ├ [...]
     └ center, centre, middle, heart, eye ("an area that is approximately central within some larger region")
        ├ inner city
        ├ financial centre
        ├ storm centre
        ├ hub
        ├ seat ("a center of authority")
        ├ [...]
        └ city centre ("the central part of a city")
```

Figure 4.1: Classification tree for city centre from WordNet 3.0. Excerpt—not all sister terms are shown.

The graphs employed for visualising the conceptual structure in Papers 2 and 3 are instances of a family of graphs used for knowledge representation called semantic networks (Sowa, 1992). Similar graphs were used in other works to design and communicate geographic ontologies. Mizen et al. (2005) used two different means of visualising what they call concept networks: network diagrams and lists of "conceptual ontology triples" where the concepts and relationships are recorded as subject-predicate-object. Both instruments were used to model geographic domain knowledge before formalising it in terms of OWL. Another variety of graphs for visualising conceptual structures are conceptual graphs (Sowa, 2000, 2008). Karalopoulos et al. (2004) presented a procedure to acquire conceptual graphs of geographic phenomena from dictionaries. Their procedure assumes that concept definitions have a determinate form and consist of a hypernym in combination with a set of differentiating statements. Figure 4.2 shows an example of a concept structure that was acquired in this way. Formally, conceptual graphs are bipartite graphs, where boxes represent concepts, and circles represent conceptual relations. The benefit of conceptual graphs is that they have formally defined semantics by relating them to common logics. However, only a limited expressiveness is provided by common logics. This led to a variety of formal and informal extensions of the standard for conceptual graphs (Sowa, 2008).

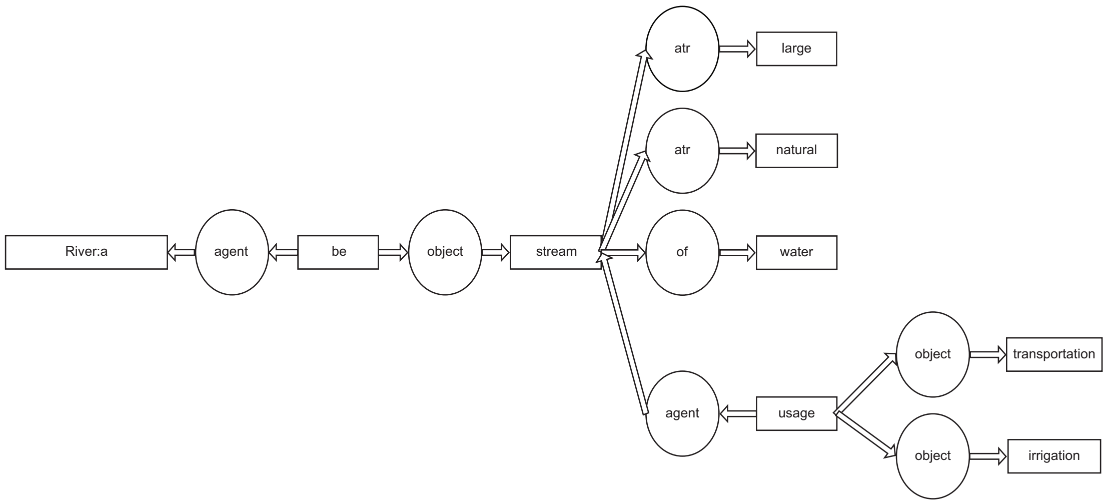


Figure 4.2: Conceptual graph for the concept „river" (Karalopoulos et al., 2004, p. 520)

(V) *To what extent is it possible to use only simple measures (such as area and topological relations) to define complex concepts?*

An aim of the work was to evaluate whether the acquired knowledge could be directly used to drive the pattern recognition process, using simple, generic analysis operations as building blocks. Modelling geographic phenomena for data enrichment proceeded in two steps both in the case study on terraced houses (Papers 2 and 3) and on city centres (Paper 4). Firstly, a conceptual model was designed based on literature analysis and an online survey, respectively. In the second step, the conceptual model was restated as a pattern recognition process, adapted to work on a set of specific datasets. The introduction of the second level is motivated by need to take into account specifics of datasets, and for formulating an efficient pattern recognition process. It can be argued that this semantic gap between symbolic modelling and algorithmic implementation lowers the level of transparency and flexibility which we aimed for. Other works in the area thus proposed to stay on a symbolic (i.e. logics) level except for calculating basic attributes. In the following, three alternative propositions are discussed with focus on how the pattern recognition process is formulated. I then present some issues encountered that complicate purely symbolic reasoning.

The approach presented by Klien (2007) and Klien and Lutz (2005) is a strategy for geospatial information discovery. In particular, a user wishes to find the dataset that most closely represents his view on a specific phenomenon (from a range of datasets available). Rather than doing mere term-based matching based on the phenomenon's name, the proposed approach consists of firstly formulating the conceptualisation in a machine-understandable way, then secondly automatically generating instances corresponding to the user's conceptualisation, and thirdly comparing the generated instances to the features available in the dataset(s). In their case study, the pattern recognition is directly carried out by an inference engine based on concept descriptions formulated in OWL or SWRL. However, they acknowledge that it may be necessary to define additional process knowledge (Klien and Lutz, 2005, p. 145).

Thomson (2009) states the hypothesis that pattern recognition can be performed based on conceptual definitions and description logics (DL) reasoning, thus eliminating the need for programming. According to Thomson, the limitations of this approach are purely technical. As DL reasoning can only handle symbolic facts, the knowledge base has to be transferred to an external application in order to carry out spatial analysis and subsequently the enriched facts have to be transferred back. She also observed limitations of current DL reasoning software when dealing with large knowledge bases and with fuzziness. This position neglects that many preparatory spatial analysis processing might be necessary for deriving the necessary symbolic facts. In Thomson's case study urban blocks were aggregated from OS MasterMap® topographic primitives by an algorithm external to the reasoning system, the ratio of different house types in each block was computed the same way, and the processing had to be triggered manually.

The motivation behind Mallenby's approach for pattern recognition (Mallenby, 2008) is to deliver user-specific representations at query-time. Third et al. (2007) extended it into a three-layered architecture to ground ontologies in data: A general layer specifying the set of existing concepts; a grounding layer which is specific to each dataset that contains the queries needed to extract the concepts from the data; finally, the data layer consists of a set of data which has been and marked up with the denotations of low level predicates such as linear, long or deep. Queries on the grounding levels are formulated in Prolog, which makes it a similar principle to the SWRL rule-based reasoning presented by Klien (2007). Hence, the necessity for implementing low level algorithms adapted to specific requirements of each concept and characteristics of the dataset remains. For example, the definition of water features (lake, river, confluence, etc.) from a topographic dataset requires complex algorithmic computations and the introduction of artificial terms to avoid errors that may be obvious to humans but not to a computer. From a practical viewpoint, a drawback might be the poor scalability of Prolog reasoning.

In the course of this work, two areas were identified that need closer attention to establish a self-contained framework of ontology-driven pattern recognition. Firstly, algorithms for specific patterns should be easy to integrate. The research addressed this issue by introducing *abstract concepts*. In Paper 3 it was shown that, without introducing an abstract concept `row of houses`, terraced houses are not recognized reliably. Hence, the potential for relying complex patterns on few, simple measures is limited. However, we have also shown that reuse of lower level patterns is possible. The city centre concept involves homogeneous residential areas, which are composed of areas of terraced housing (along with detached and semi-detached housing).

The second area concerns workflow management. For a complete pattern recognition process, there are preparatory and bookkeeping operations that do not directly relate to the concept definition. Examples for such pre-processing operations are presented in Paper 4. A Points of Interest dataset was generated by integrating OS MasterMap® Address Layer 2 and OS Points of Interest. For generating industrial sites, the Points of Interest dataset was matched to buildings from OS MasterMap® Topography Layer. By introducing a graphical workflow management for the definition of the workflow process, it would be possible to retain some of the transparency and flexibility, while keeping the efficiency of algorithmic solutions. The thesis did not dwell on matters of workflow management, however, some ideas are sketched in the Outlook.

4.1.4 Investigating the role of uncertainty in the ontology-driven data enrichment approach

(VI) *How can we integrate vagueness into the data enrichment process?*

To conduct pattern recognition, vague predicates in concept definitions, such as ‘small area’ or ‘close to’, have to be translated to numerical values. Section 2.2.4 introduced the different stances that were taken in order to deal with such kinds of vagueness. To deal with vague predicates, Mizen et al. (2005) and Klien (2008) employ scales of categories for vague terms that can be mapped to rational numbers. Klien (2008) coins the notion *reference spaces* to denote these mappings (Figure 4.3). Mallenby (2008) applies the same principle, though he takes a supervaluationist stance and studies the influence of the choice of rational numbers to the regions produced by the pattern recognition procedure. The benefit of such approaches is that means for classical logic reasoning can be employed. However, since these mappings are context-dependent, they have to be defined by a domain expert for each individual case.

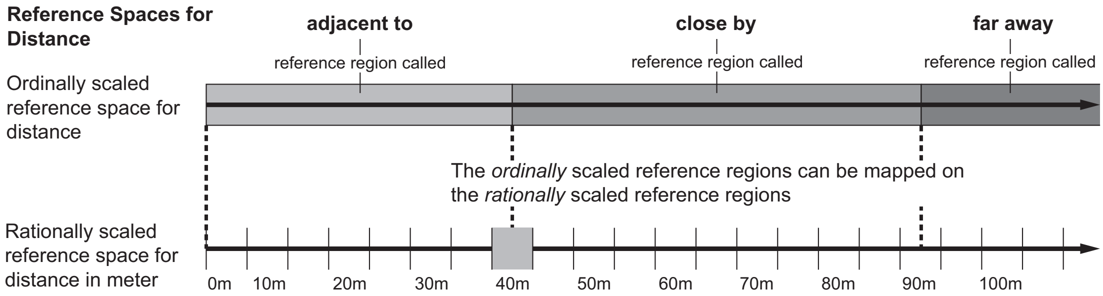


Figure 4.3: Exemplary reference spaces for distance (Klien, 2008, p. 134)

This thesis approached vagueness by means of graded membership. Paper 3 demonstrated the potential of combining semantic modelling with probabilistic reasoning. The meaning of predicates is given by a probability distribution. The probability distributions are context-dependent; however it is possible to use machine learning for estimating them. In Paper 4, a human subject experiment that aimed at judging city centre typicality from panorama situations was conducted. The experiment suggests that being inside or outside of a city centre exhibits a vague nature. The thematic vagueness of city centre typicality results in a spatial vagueness of locating boundaries. Additionally, the city centre concept exhibits definitional vagueness; it is easy to point to an exemplar of city centre, a definition of the concept appears a difficult problem. Hence, human subject experiments were conducted to define a common denominator of the concept.

4.2 Evaluation of the ontology-driven methodology

In the following, strengths and limitations of the developed methodology for ontology-driven pattern recognition are examined.

4.2.1 Strengths

Separation of domain knowledge from pattern recognition process. Improved representation of ontological assumptions behind the abstraction of geographical phenomena is a major motivation for the development of ontology-driven pattern recognition. Explicitly modelling geographic phenomena as suggested in this thesis improves an increased level of transparency and offers many benefits for information integration. The conceptual models allow for communicating with domain experts whether the conceptualisation behind the pattern recognition process is correct. They can be used by customers to evaluate the fitness-for-use of a representation. If formalised in a language such as OWL, explicit models can be used for automated discovery of geographic information. Ideally, an ontology-driven approach to pattern recognition allows for improved flexibility to adapt concepts to specific needs and different datasets. This is limited somewhat due to specific implementations needed for a conceptualisation (see Section 4.1.3). However a way to improve flexibility of algorithmic implementation is sketched in Section 5.3.1.

Reduction of conceptual bias. Map generalisation aims at improving fitness-for-use by adaptation of representations to specific needs. The top-down perspective taken in this research ensures that the abstraction is guided by phenomenological knowledge. This leads to a smaller conceptual bias between human conceptualisation and what is represented in the spatial database.

Uncertainty of geospatial phenomena. As was seen in the Literature Review, there are various types of uncertainty adhering to geospatial phenomena. The research accounts for vagueness by adopting a model of graduated membership. It also demonstrates how definitional uncertainty can be addressed by human subject experiments.

Combining structural knowledge with probabilistic reasoning. The approach showed in Paper 3 combines syntactical and statistical approaches to pattern recognition (Jain et al., 2000). Thus, it allows a domain expert to model structural knowledge about geographic phenomena explicitly, while using machine learning to obtain knowledge about vague thresholds.

Tested on extensive datasets. Previous approaches for ontology-driven pattern recognition were either discussed on a conceptual level, or were conducted on small areas only. The two-stage approach presented in this thesis was tested successfully on large extracts of commercially available datasets. It was hence argued that practicability demands a trade-off between symbolic knowledge representation and computationally efficiency.

4.2.2 Limitations and open problems

Potential of available datasets. The success of discovering higher level semantics is dependent on the datasets that are available. For example, the potential to recover land use from land cover alone is limited since some types of land use yield similar spatial configurations. A group of large buildings may be part of an industrial complex, of a commercial district, or it may be large residential blocks. However, it is possible to recover land use by integrating land cover and gazetteer data. A limitation of this type was observed for the concept of old town, which requires that building age is represented; likewise, for reliable detection of retail parks and business centres, building height would be valuable. It is envisaged that data enrichment from multiple integrated datasets gains importance as it becomes easier (and cheaper) to access data. A direct application in the context of the thesis is using building heights. They can be obtained automatically through laser scanning or radar interferometry (e.g. Gamba & Houshmand, 2000).

Identification of vernacular places. This issue is connected to the limitations of explanatory power of datasets. Paper 4 dwelled on definition and recognition of a concept that has, at least to some extent, a place-like character (Cresswell, 2004). It has to be noted here that there are factors beyond physical and economical structure that dictate place identification. Davies (2009) observes that vernacular places such as urban neighbourhoods occasionally expand over areas featuring a diversity of functional units and urban styles. She lists supplementary identifying factors for place identification: Social coherence, individual "home range" of people, local social or political activity, and media coverage, to name a few. Meanings of such vernacular place names cannot be captured by topographic datasets, but by studying how people use them. Traditional methods to capture an agreed extent of such areas involve interviewing locals (e.g. Campbell et al., 2009) and are accordingly laborious. This might be migitated by using georeferenced information on the web as a data source (Hollenstein & Purves, 2010; Jones et al, 2008), however attention needs to be paid to the representativity of data obtained in this way (Hollenstein & Purves, 2010).

Integration of knowledge modelling and pattern recognition process. It was criticised that the execution of pattern recognition external to a system where the conceptual model is formally represented does not lead to a well integrated system (Thomson, 2009, pp. 223−224). Indeed it could be argued that the introduction of a symbolic grounding layer, as presented by Third et al. (2007), is a better integrated way of formulating pattern recognition knowledge. However it has to be noted that there is a close relation between the grounding layer and algorithmic structure recognition algorithms in their framework too, such that one is not delivered of algorithmic programming when one changes the conceptualisation or uses another dataset.

Chapter 5

Conclusions

This thesis presented research about enriching topographic datasets with higher order phenomena and thus adapting general representations to very specific uses. The main motivation was to develop a top-down methodology that is driven by phenomena's semantics. Departing from this general aim, the following six *research questions* were addressed:

(I) How can semantic modelling help in the development of cartographic pattern recognition methods?

(II) What are the requirements for an ontology-driven approach to data enrichment in an urban context?

(III) What methods are available for extracting knowledge about urban structures?

(IV) Can urban structures be decomposed in terms of the phenomenological approach?

(V) To what extent is it possible to use only simple measures (such as area and topological relations) to define complex concepts?

(VI) How can we integrate vagueness into the data enrichment process?

This concluding chapter highlights the main contributions and insights of the thesis, and provides an outlook on future developments.

5.1 Main contributions

This thesis established a new perspective on cartographic pattern recognition by adopting a top-down methodology. Two case studies, each modelling a specific urban structure, were conducted to demonstrate the methodology. Each case study aimed to look at a pattern recognition process in a holistic way, i.e. from knowledge acquisition, execution of pattern recognition using large-scale topographic vector data, to evaluation of the output using comparative sources. With respect to the research objectives set out above, the following contributions were made:

- The thesis discussed the challenges for producers of topographic datasets in meeting user requirements for very specific representations. It identified the benefits of an ontology-driven approach to pattern recognition in comparison to purely algorithmic approaches (Paper 1). These are provision of better flexibility and increased transparency.
- A two-level approach was proposed to incorporate semantics into the pattern recognition workflow (Paper 2 and Paper 4). At first, domain knowledge is explicitly captured in conceptual models. These models explain a phenomenon in terms of its geometrical properties and spatial relations. The models are subsequently used to inform the pattern recognition process.
- The research employed a phenomenological approach to decompose knowledge of urban structures by relating them to other, possibly simpler concepts and measures. It was shown that employing only simple, generic measures is limited due to the complexity of the pattern recognition task. Hence a component-oriented approach to incorporate specific algorithms was proposed.
- A model of vagueness based on graded membership is adopted throughout the research. This model adheres to observations that people judge some exemplars to be more typical instances of a class than others.
- The research developed a methodology to integrate semantic modelling with Bayesian inference to carry out pattern recognition (Paper 3). The methodology is a means to overcome difficulties with the vague nature of terms that describe geographic phenomena in conceptual models. Using Bayesian inference allows learning the influence of predicates on classification results from training data.
- An online survey was conducted for acquiring the meaning of 'city centre' (Paper 4). The outcomes suggest that human subject experiments are a reasonable means to capture human conceptualisations of complex geographic phenomena. The survey provided firstly information to define a city centre, and secondly comparative values for verifying model outputs were obtained in an experiment that requested the participants to judge city centre typicality from panoramic images.

5.2 Insights

The research in this thesis was conducted in an iterative process of conceptual work, prototype implementation and experimenting with real datasets. This section summarises crucial insights gained in the course of this work.

Ongoing need for generalisation of topographic information. As current systems are able to store and process ever larger quantities of data, and paper maps are no longer the primary medium to portray geographic information, one might postulate that map generalisation methods have become redundant. However, it is quite contrary. Ongoing efforts to integrate geographic datasets and make them better accessible and the proliferating use of geographic information in various disciplines require mediation between different conceptualisations. However, there are needs for more explicit semantics and flexibility in generalisation methods to accommodate for very specific uses.

Diversity of urban structures. A challenge for urban structure recognition is the wealth of structures that exist. The forming of urban structures is subject to cultural context, customs of individual building periods, and history and geographic setting of individual cities. The richness of urban form makes it difficult to develop universally valid data enrichment procedures. While this is a motivation to develop more transparent approaches, it also means the developed procedures should be tested extensively to reveal their applicability and limitations.

Urban space is imbued with social meaning. Unlike the mountains, hills, and valleys of the physical environment, urban space is most predominantly an artificially shaped space. It is formed by social interactions and it influences them in turn (Hillier & Hanson, 1984). While map generalisation research was concerned with geometrical optimisations and aesthetic quality for a great deal, it would definitely benefit of having a closer look on the meanings of place, and how it could be formally modelled.

Inference method. This research evolved concurrently with other works in the geographic domain which can be commonly placed under the umbrella "ontology-driven pattern recognition" (i.e., Klien, 2008; Mallenby, 2008; Thomson, 2009). This work used custom algorithms to carry out pattern recognition rather than employing logics-based reasoning. The decision towards custom algorithms was made due to current technological limitations of logical reasoning engines to deal with vagueness and large quantities of data, and the tight interrelation between pattern recognition process and low level measures. Thus, a division into conceptual modelling and algorithmic implementation seems sensible.

5.3 Outlook

5.3.1 Suggested improvements and future developments

5.3.1.1 Composition of complete data enrichment workflows

As argued in Section 4.1.3, it seems currently not practical to compile all parts of a data enrichment process automatically from conceptual definitions of phenomena. However, to keep a certain level of transparency and flexibility, it is beneficiary to model such processes explicitly, i.e. by means of workflows which can be graphically designed and altered. Petzold et al. (2006) described the employment of workflow management systems for orchestrating automated generalisation operations. The research challenge to be addressed is to develop procedures to ensure consistency between conceptual description and pattern recognition workflow.

Such a workflow management system would offer an extendible library of algorithms for basic measures and available abstract concepts to be embedded into a workflow. This requires that an ontology is developed to describe capabilities and context of each algorithm (cf. Regnauld, 2007). A web service architecture (Neun, 2007) might be used to integrate algorithms into the workflow management system. A further research challenge, which was also put forth by Steiniger (2007) and Mallenby (2007), is to systematically assess basic algorithms for their applicability and generality to be used in different contexts.

5.3.1.2 Development of a comprehensive system for data enrichment

As versatile data enrichment is shifting from research interest to business requirement (Parker, 2004), there is a need to build user-friendly systems that integrate the complete range of concept discovery, concept and workflow definition, execution of data enrichment, and storage of enriched data. In the following, open issues of user interaction and storage will be discussed.

Design of user interaction schemes: A complete system would also require the design of user interaction schemes. As the system should be operated by domain expert, interaction schemes need to be found that guide a user while defining new concepts (e.g., using concept maps) and for translating conceptual definitions into a pattern recognition procedure (e.g., using a workflow engine as sketched above). A research opportunity is to develop methods to visualise an enriched database, including visualisation of uncertainty and visualisation of generated relation instances for an entity. A second area that needs closer attention is the design of interaction schemes to browse through the concepts of an enriched database, grasp the meaning of concepts and decide upon fitness-for-purpose. Gahegan and Pike (2006) show the complexity involved in designing such schemes.

Storage of enriched data: This thesis did not dwell into issues of storing the enriched data. As real-time enrichment is too time-consuming in many cases, the produced entities should be stored in a database for later retrieval. In doing so it is beneficiary to model references to composing entities, thus creating a multiple-representation database (Kilpeläinen, 1997; Burghardt et al., 2010). Two issues are interesting for further research in this context. Firstly, if a component is updated, for instance a building is demolished, related higher level concepts need to be reprocessed. Thereby also neighbourhood effects are to be considered. For instance, if a derelict industrial site is developed into a shopping centre, it might influence the boundary of the city centre. Such neighbourhood effects are a challenge in map generalisation as well (Touya, 2010). The second issue is how objects with vague boundaries such as city centres can be represented in a database.

5.3.1.3 Extension to 3-D and time

The concentration of activities brings about a vertical layering of functions in urban spaces. Hence, there is a need for topographic information in three dimensions which is recognised by data producers (Stoter & Salzmann, 2003). The Swiss national mapping agency Swisstopo is currently releasing their new product TLM, which comprises an accurate and three dimensional representation of Switzerland's physical environment, including heights of man-made constructions and building roof structures (O'Sullivan et al., 2008). Methods are developed that allow automatically capturing building façade structures from terrestrial laser scanning data (e.g. Pu, 2008). This will make it possible to capture highly detailed city models at low cost. CityGML (Kolbe et al., 2005) was devised to store and exchange such models in various levels of details. The widespread availability of 3-D urban models opens up opportunities to carry out large-scale analysis of urban character that were previously impractical due to the effort for carrying out extensive ground surveys. The main challenge will be to find efficient methods for processing large quantities of 3-D data.

As map producers are shifting to digital, vector-based production lines, historical states of settlements are available, both of physical and functional nature. This makes it possible to analyse not only urban configurations, but also urban processes. Hence, it would be interesting to investigate whether processes can be formalised and integrated into the framework in the same way as urban configurations, and how process knowledge can be linked to sequences of urban configurations.

5.3.2 Final thoughts

> *"You can know the name of a bird in all the languages of the world, but when you're finished, you'll know absolutely nothing whatever about the bird...So let's look at the bird and see what it's doing – that's what counts."* (Richard P. Feynman, 1966)[1]

This thesis took a modelling perspective to spatial data enrichment and argued that it ought to respect and commence with the meaning of geographic phenomena. This is by no means a novel claim (Nyerges, 1991), however to date it gained relatively little attention in the map generalisation community. Rather than attempting to mimic cartographers in the design of a general purpose map, this thesis understood the map generalisation process as adaptation of general representations to specific contexts where geographic information is used (whether this context is a professional area or is part of common geographic experience) and indicated implications to the design of pattern recognition processes. Semantic enrichment is an indispensable tool for meeting specific requirements while analysing, integrating, and visualising topographic information. It is hoped that this thesis brings forward the understanding of conceptual abstraction in map generalisation and is a contribution towards improved versatility of geographic information.

[1] Richard Feynman used this statement at the fifteenth annual meeting of the National Science Teachers Association in New York City in a talk titled "What is Science?". It was recovered by Gahegan and Pike (2006), and gratefully adopted by me.

Bibliography

Agarwal, P. (2004). Contested Nature of Place: Knowledge Mapping for Resolving Ontological Distinctions Between Geographical Concepts. In M. J. Egenhofer, C. Freksa, & H. J. Miller (Eds.), *Geographic Information Science, Third International Conference, GIScience 2004*. Lecture Notes in Computer Science, Vol. 3234 (pp.1–21). Berlin / Heidelberg: Springer-Verlag.

Agarwal, P. (2005a). Ontological considerations in GIScience. *International Journal of Geographical Information Science, 19*(5), 501–536.

Agarwal, P. (2005b). Topological and Geometric Operators for Ontological Classification of Vague Geographic Concepts. *Spatial Cognition & Computation, 5*(1), 69–88.

Anders, F., née Heinzle (2007). Mustererkennung in Straßennetzwerken. Verfahren zur Interpretation von Vektordaten. (Doctoral dissertation). Leibniz Universität Hannover, Hannover. München: Verlag der Bayerischen Akademie der Wissenschaften.

Anders, K.-H., Sester, M., & Fritsch, D. (1999). Analysis of Settlement Structures by Graph-Based Clustering. In W. Förstner, C.-E. Liedtke, & J. Bückner (Eds.), *SMATI'99. Semantic Modelling for the Acquisition of Topographic Information from Images and Maps* (pp. 41–49). ftp://ftp.tnt.uni-hannover.de/pub/smati99/proceeding.pdf. Accessed 26.11.2010.

Armstrong, M. P. (1991). Knowledge classification and organization. In B. P. Buttenfield & R. B. McMaster (Eds.), *Map Generalization: Making Rules for Knowledge Representation* (pp. 59–85). Essex: Longman Scientific and Technical.

Baccini, P., Baumgartner, F., Lichtensteiger, T., Michaeli, M., & Thalmann, E. (2007). Urbane Schweiz. In OcCC/ProClim–. (Eds.), *Klimaänderung und die Schweiz 2050. Erwartete Auswirkungen auf Umwelt, Gesellschaft und Wirtschaft* (pp. 123–136). Bern: OcCC.

Bafna, S. (2003). Space Syntax. A Brief Introduction to Its Logic and Analytical Techniques. *Environment and Behaviour, 35*(1), 17–29.

Barnsley, M. J., & Barr, S. L. (1997). Distinguishing Urban Land-use Categories in Fine Spatial Resolution Land-cover Data using a Graph-based, Structural Pattern Recognition System. *Computers, Environment and Urban Systems, 21*(3–4), 209–225.

Barr, S. L., Barnsley, M. J., & Steel, A. (2004). On the separability of urban land-use categories in fine spatial scale land-cover data using structural pattern recognition. *Environment and Planning B, 31*(3), 397–418.

Batty, M. (2005). *Cities and Complexity. Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals.* Cambridge, USA: MIT Press.

Benenson, I., & Torrens, P. M. (2005). *Geosimulation. Automata-based modeling of urban phenomena.* Chichester, UK: John Wiley & Sons Ltd.

Bennet, B. (2001). What is a Forest? On the Vagueness of Certain Geographic Concepts. *Topoi 20*(2), 189–201.

Bennet, B. (2010). Spatial Vagueness. In R. Jeansoulin, O. Papini, H. Prade, & S. Schockaert (Eds.), *Methods for Handling Imperfect Spatial Information.* Studies in Fuzziness and Soft Computing, Vol. 256 (pp. 15–47). Berlin / Heidelberg: Springer-Verlag.

Bennett, B., & Agarwal, P. (2007). Semantic categories underlying the meaning of 'place'. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers, (Eds.), *Spatial Information Theory: 8th International Conference, COSIT 2007.* Lecture Notes in Computer Science, Vol. 4736 (pp. 78–95). Berlin / Heidelberg: Springer-Verlag.

Bennett, B., Mallenby, D., & Third, A. (2008). An Ontology for Grounding Vague Geographic Terms. In C. Eschenbach & M. Grüninger (Eds.), *Formal Ontology in Information Systems – Proceedings of the Fifth International Conference. FOIS-08* (pp. 280–293). Amsterdam: IOS Press.

Bertin, J. (1967, 1999). *Sémiologie graphique: Les diagrammes – Les réseaux – Les cartes* (Original edition published 1967 by Gauthier-Villars, Paris). Paris: Les réimpressions des Éditions de l'École des Hautes Études en Sciences Sociales.

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science, 12*(4), 299–314.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(1), 2–16.

Boffett, A. (2001). Méthode de création d'information multi-niveaux pour la généralisation cartographique de l'urbain. (Unpublished doctoral dissertation). Université de Marne-la-Vallée, Marne-la-Vallée.

Boffet, A., & Rocca Serra, S. (2001). Identification of spatial structures within urban blocks for town characterization. *Proceedings 20th International Cartographic Conference* (pp. 1974–1983). http://icaci.org/documents/ICC_proceedings/ICC2001/. Accessed 02.12.2010.

Boffet, A., & Coquerel, C. (2000). Urban Classification for Generalization Orchestration. In D. Fritsch & M. Molenaar (Eds.), *19th Congress of the International Society for Photogrammetry and Remote Sensing.* International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B4 (pp. 132–139). http://www.isprs.org/proceedings/XXXIII/congress/part4/default.aspx. Accessed 29.11.2010.

Brassel, K. E., & Weibel, R. (1988). A Review and Framework of Automated Map Generalization. *International Journal of Geographical Information Systems, 2*(3), 229–244.

Bromley, R. D. F., Tallon, A. R., & Thomas, C. J. (2003). Disaggregating the space-time layers of city-centre activities and their users. *Environment and Planning A, 35*(10), 1831–1851.

Bundesamt für Raumentwicklung ARE (2009). *Monitoring urbaner Raum Schweiz – Analysen zu Städten und Agglomerationen.* www.bundespublikationen.admin.ch, Art. Nr. 812.071.d.

Burghardt, D., Petzold, I., & Bobzien, M. (2010). Relation Modelling within Multiple Representation Databases and Generalisation Services. *The Cartographic Journal, 47*(3), 238–249.

Burnett, C., & Blaschke, T. (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling, 168*(3), 233–249.

Burrough, P. A. (1989). Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science, 40*(3), 477–492.

Campbell, E., Henly, J. R., Elliott, D. S., Irwin, K. (2009). Subjective Constructions of Neighbourhood Boundaries: Lessons from a Qualitative Study of Four Neighourhoods. *Journal of Urban Affairs, 31*(4), 461–490.

Carter, H. (1995). *The Study of Urban Geography* (4th ed.). London: Arnold.

Chaudhry, O. Z. (2007). Modelling Geographic Phenomena at Multiple Levels of Detail: A Model Generalisation Approach Based on Aggregation. (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh.

Chaudhry, O. Z., & Mackaness, W. A. (2006). Creation of Fiat Boundaries in Higher order Phenomenon. *9th Workshop of the ICA Commission on Generalisation and Multiple Representation.* http://aci.ign.fr/Portland/program.php. Accessed 01.12.2010.

Chaudhry, O. Z., & Mackaness, W. A. (2008). Automatic identification of urban settlement boundaries for multiple representation databases. *Computers, Environment and Urban Systems, 32*(2), 95–109.

Chaudhry, O. Z., Mackaness, W. A., & Regnauld, N. (2009). A functional perspective on map generalisation. *Computers, Environment and Urban Systems, 33*(5), 349–362.

Christophe, S., & Ruas, A. (2002). Detecting Building Alignments for Generalisation Purposes. *Proceedings of the Symposium on Geospatial Theory, Processing and Applications.* http://www.isprs.org/proceedings/XXXIV/part4/paper.html. Accessed 25.11.2010.

Cohn, A. G., & Gotts, N. M. (1996). The ‘Egg-Yolk’ Representation of Regions with Indeterminate Boundaries. In P. Burrough & A. Frank (Eds.), *Geographic Objects With Indeterminate Boundaries* (pp.171–187). London: Taylor & Francis.

Conzen, M. R. G. (1960, 1969). *Alnwick, Northumberland. A study in town-plan analysis*. Institute of British Geographers, Publication No. 27 (2nd ed. Original edition published 1960 by George Philip, London). London: Institute of British Geographers.

Couclelis. H. (1992). Location, place, region, and space. In R. F. Abler, M. G. Marcus, & J. M. Olson (Eds.), *Geography's Inner Worlds. Pervasive Themes in Contemporary American Geography.* New Brunswick, NJ: Rutgers University Press.

Couclelis, H. (1996). Towards an operational typology of geographic entities with ill-defined boundaries. In P. Burrough & A. Frank (Eds.), *Geographic Objects With Indeterminate Boundaries* (pp. 45–55). London: Taylor & Francis.

Couclelis, H. (2009). The Abduction of Geographic Information Science: Transporting Spatial Reasoning to the Realm of Purpose and Design. In K. S. Hornsby, C. Claramunt, M. Dennis, & G. Ligozat (Eds.), *Spatial Information Theory, 9th International Conference, COSIT 2009.* Lecture Notes in Computer Science, Vol. 5756 (pp. 342–356). Berlin / Heidelberg: Springer-Verlag.

Cresswell, T. (2004). *Place: a short introduction.* Malden, MA: Blackwell Publishing Ltd.

Crevier, D., & Lepage, R. (1997). Knowledge-Based Image Understanding Systems: A Survey. *Computer Vision and Image Understanding, 67*(2), 161–185.

Davies, C. (2009). Are Places Concepts? Familiarity and Expertise Effects in Neighborhood Cognition. In K. S. Hornsby, C. Claramunt, M. Dennis, & G. Ligozat (Eds.), *Spatial Information Theory, 9th International Conference, COSIT 2009.* Lecture Notes in Computer Science, Vol. 5756 (pp. 36–50). Berlin / Heidelberg: Springer-Verlag.

Davies, C., Holt, I., Green, J., Harding, J., & Diamond L. (2009). User Needs and Implications for Modelling Vague Named Places. *Spatial Cognition & Computation, 9*(3), 174–194.

Davies, C., Wood, L., & Fountain, L. (2005). User-centred GI: hearing the voice of the customer. *Annual conference of the Association for Geographic Information AGI'05, People, Places and Partnerships*. http://www.ordnancesurvey.co.uk/oswebsite/partnerships/research/publications/docs/2005/ClareDavies etal geo.pdf. Accessed 25.01.2011.

De Man, W. H. E. (2006). Understanding SDI; complexity and institutionalization. *International Journal of Geographical Information Science, 20*(3), 329–343.

Dehbi, Y., Plümer, L. (2011). Learning grammar rules of building parts from precise models and noisy observations. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(2), 166–176.

Douglas, D. & Peucker, T. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer 10*(2), 112–122.

Earl, D. (2010). *Vague Analysis. Metaphysica, 11*(2), 223–233.

Edwardes, A. (2007). Re-placing Location: Geographic Perspectives in Location Based Services. (Unpublished doctoral dissertation). University of Zurich, Zurich.

Egenhofer, M. J., & Herring, J. R. (1990). A Mathematical Framework for the Definition of Topological Relationships. *Proceedings of the 4th International Symposium on Spatial Data Handling* (pp. 803–813). Zurich: University of Zurich.

Egenhofer, M. J., & Mark, D. M. (1995). Naïve Geography. In A. U. Frank & W. Kuhn (Eds.), *Spatial Information Theory. A Theoretical Basis for GIS. Proceedings International Conference COSIT '95*. Lecture Notes in Computer Science, Vol. 988 (pp. 21–23). Berlin / Heidelberg: Springer-Verlag.

Ellenkamp, Y., & Maessen, B. (2009). Napoleon's registration principles in present times: The Dutch System of Key Registers. *GSDI 11 World Conference, Workshop 4.5 (SDI in National Government: Dutch Key GeoRegisters)*. http://www.gsdi.org/gsdiconf/gsdi11/. Accessed 04.09.2011.

Fisher, P. F. (1999). Models of uncertainty in spatial data. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical Information Systems: Principles and technical issues*.(2nd ed., Vol. 1, pp. 191–205). New York: John Wiley & Sons.

Fisher, P. F. (2000a). Sorites paradox and vague geographies. *Fuzzy Sets and Systems, 113*(1), 7–18.

Fisher, P. F. (2000b). Fuzzy modeling. In S. Openshaw, R. Abrahart, & T. Harris (Eds.), *Geocomputing* (pp. 161–186). London: Taylor & Francis.

Fisher, P. F., Wood, J., & Cheng, T. (2004). Where is Helvellyn? Multiscale morphometry and the mountains of the English Lake District. *Transactions of the Institute of British Geographers, 29*(1), 106–128.

Gahegan, M., & Pike, W. (2006). A Situated Knowledge Representation of Geographical Information. *Transactions in GIS, 10*(5), 727–749.

Gamba, P., & Houshmand, B. (2000). Digital Surface Models and Building Extraction: A Comparison of IFSAR and LIDAR Data. *IEEE Transactions on Geoscience and Remote Sensing, 38*(4), 1959–1968.

Gärdenfors, P. (1996). Conceptual spaces as a basis for cognitive semantics. In A. Clark, J. Ezquerro, & J. M. Larrazabal (Eds.), *Philosophy and cognitive science: categories, consciousness, and reasoning. Proceedings of the Second International Symposium on Cognitive Science* (pp. 159–180). Dordrecht: Kluwer.

Gartner, G. (2004). Location-based mobile pedestrian navigation services— the role of multimedia cartography. *International Joint Workshop on Ubiquitous, Pervasive and Internet Mapping (UPIMap2004).* http://www.ubimap.net/upimap2004/. Accessed 01.02.2011.

Greenwood, J., & Mackaness, W. A. (2002). Revealing Associative Relationships for Complex-Object Creation. In M. J. Egenhofer & D. M. Mark (Eds.), *GIScience 2002 Abstracts. The Second International Conference on GIS* (pp. 48–55). University of California Regents.

Grünreich, D. (1985). Computer-assisted generalisation. *Papers CERCO Cartography Course.* Frankfurt am Main: Institut für Angewandte Geodäsie.

Goldstein, L. (2000). *How to boil a live frog. Analysis, 60*(2), 170–178.

Griffiths, S., Jones, C. E., Vaughan, L., & Haklay, M. (2010). The persistence of suburban centres in Greater London: combining Conzenian and space syntax approaches. *Urban Morphology, 14*(2), 85–99.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies, 43*(5–6), 907–928.

Guarino, N., & Giaretta, P. (1995). Ontologies and Knowledge Bases – Towards a Terminological Clarification. In N. Mars (Ed.*), Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing* (pp. 25–32). Amsterdam: IOS Press.

Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino (Ed.), *Formal Ontology in Information Systems. Proceedings of FOIS'98* (pp. 3–15). Amsterdam: IOS Press.

Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Systems, 17*(3), 235–249.

Haggett, P. (2001). *Geography. A Global Synthesis.* Harlow, UK: Prentice Hall.

Haggett, P., & Chorley, R. J. (1967). Models, Paradigms, and the New Geography. In R. J. Chorley & P. Haggett (Eds.), *Models in geography* (pp. 19–41). London: Methuen.

Hart, G., & Greenwood, J. (2003). A Component Based Approach to Geo-Ontologies and Geodata Modelling to Enable Data Sharing. In M. Gould, R. Laurini, & S. Coulondre (Eds.), *AGILE 2003, 6th AGILE Conference on Geographic Information Science* (pp. 197–206). Lausanne: Presses polytechniques et universitaires romandes.

Harvey, F. (1997). Generalization and GIS - Holism and Purpose, Conceptualization and Representation. *ICA Workshop on Map Generalization*. http://www.tc.umn.edu/~fharvey/research/gengeomatch/gi_generalization.html. Accessed 22.11.2010.

Heinzle, F., & Anders, K.-H. (2007). Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks. In W. Mackaness, A. Ruas, & T. Sarjakoski (Eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications* (pp. 233–253). Amsterdam: Elsevier.

Hillier, B., & Hanson, J. (1984). *The social logic of space.* Cambridge: Cambridge University Press.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., & Xu, J. (1993). Natural movement; or, configuration and attraction in urban space use. *Environment and Planning B, 20*(1), 29–66.

Hollenstein, L., & Purves, R. S. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, No 1(2010), 21–48.

ICA (International Cartographic Association). (1973). *Multilingual Dictionary of Technical Terms in Cartography.* Wiesbaden: Franz Steiner Verlag.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(1), 4–37.

Janowicz, K., Adams, B., & Raubal, M. (2010). Semantic Referencing – Determining Context Weights for Similarity Measurement. In S. I. Fabrikant, T. Reichenbacher, M. van Kreveld, & C. Schlieder (Eds.), *Geographic Information Science: Sixth International Conference, GIScience 2010*. Lecture Notes in Computer Science, Vol. 6292 (pp. 70–84). Berlin / Heidelberg: Springer-Verlag.

Jones, C. B., Purves, R. S., Clough, P. D., & Joho, H. (2008). Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science, 22*(10), 1045–1065.

Jones P., Patterson, J., & Lannon, S. (2007). Modelling the built environment at an urban scale—Energy and health impacts in relation to housing. *Landscape and Urban Planning, 83*(1), 39–49.

Joubran, J., & Gabay, Y. A. (2000). Method for Construction of 2D Hull for Generalized Cartographic Representation. In D. Fritsch & M. Molenaar (Eds.), *19th Congress of the International Society for Photogrammetry and Remote Sensing.* International Archives of Photogrammetry and Remote Sensing. IAPRS Vol. XXXIII, Part B4 (pp. 417–424). http://www.isprs.org/proceedings/XXXIII/congress/part4/default.aspx. Accessed 29.11.2010.

Kaplan, D., Wheeler, J., & Holloway, S. (2009). *Urban Geography* (2nd ed.). New York: John Wiley & Sons.

Karalopoulos, A., Kokla, M., & Kavouras, M. (2004). Geographic Knowledge Representation Using Conceptual Graphs. *7th AGILE Conference on Geographic Information Science*, 29. April–1. May, 2004, Heraklion, Greece. http://plone.itc.nl/agile_old/Conference/greece2004/heraklion.html. Accessed 10.02.2011.

Kavouras, M., & Kokla, M. (2002). A method for the formalization and integration of geographical categorizations. *International Journal of Geographical Information Science, 16*(5), 439–453.

Kilpeläinen, T. (1997). Multiple representation and generalization of geo-databases for topographic maps. (Doctoral dissertation). Publications of the Finnish Geodetic Institute, No. 124, Kirkkonummi.

Klien, E. (2008). Semantic Annotation of Geographic Information. (Unpublished doctoral dissertation). Westfälische Wilhelms-Universität Münster, Münster.

Klien, E. (2007). A Rule-Based Strategy for the Semantic Annotation of Geodata. *Transactions in GIS, 11*(3), 437–452.

Klien, E., & Lutz, M. (2005). The Role of Spatial Relations in Automating the Semantic Annotation of Geodata. In A. G. Cohn & D. M. Mark (Eds.), *Spatial Information Theory, International Conference, COSIT 2005.* Lecture Notes in Computer Science, Vol. 3693 (pp. 133–148). Berlin / Heidelberg: Springer-Verlag.

Klippel, A., Hansen, S., Richter, K.-F., & Winter, S. (2009). Urban granularities—a data structure for cognitively ergonomic route directions. *GeoInformatica 13*(2), 223–247.

Kolbe, T. H., Gröger, G., & Plümer, L. (2005). CityGML – Interoperable Access to 3D City Models. In P. van Oosterom, S. Zlatanova, & E. Fendel (Eds.), *Geo-information for Disaster Management* (pp. 883–899). Berlin / Heidelberg: Springer-Verlag.

Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science, 15*(7), 613–631.

Kuhn, W. (2004). *Why Information Science needs Cognitive Semantics - and what it has to offer in return.* Version 2.0. http://ifgiweb.uni-muenster.de/~kuhn/research/publications/pdfs/discussion%20papers/Why%20Cognitive%20Semantics.pdf. Accessed 25.11.2010.

Kulik, L. (2001). A Geometric Theory of Vague Boundaries Based on Supervaluation. In D. R. Montello (Ed.), *Spatial Information Theory. Foundations of Geographic Information Science, COSIT 2001*. Lecture Notes in Computer Science, Vol. 2205 (pp. 44–59). Berlin / Heidelberg: Springer-Verlag.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* Chicago, IL: The University of Chicago Press.

Lannon, S., & Linovski, O. (2009). Ontologies for the Classification of Urban Characteristics: Opportunities for Urban Designers. In J. Teller, A.-F. Cutting-Decelle, & R. Billen (Eds.), *Future of Urban Ontologies. Proceedings of the Final Conference of the COST Action C21 –Towntology* (pp. 59–67). Liège: Les éditions de l'Université de Liège.

Larson, R. R. (1996). Geographic information retrieval and spatial browsing. In L. C. Smith & M. Gluck (Eds.), *Geographic Information Systems and Libraries: Patrons, Maps, and Spatial Information. Proceedings of the 32nd Clinic on Library Applications of Data Processing* (pp. 81–123). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Lawton, R., & Pooley, C. (1976). *The social geography of Merseyside in the nineteenth century.* Final Report to the Social Science Research Council. London: SSRC.

Liu, Y., Molenaar, M., Ai, T., & Liu, Y. (2003). Categorical Database Generalization aided by Data Model. *Proceedings of the 21st International Cartographic Conference (ICC)* (CD-ROM, pp. 2308–2318). http://www.itc.nl/library/papers_2003/art_proc/liu.pdf. Accessed 28.01.2011.

Luck, M., & Wu, J. (2002). A gradient analysis of urban landscape pattern: a case study from the Phoenix metropolitan region, Arizona, USA. *Landscape Ecology, 17*(4), 327–339.

Lynch, K. (1960). *The Image of the City.* Cambridge, MA: MIT Press.

Lüscher, P., Burghardt, D., & Weibel, R. (2007). Ontology-driven Enrichment of Spatial Databases. *10th ICA Workshop on Generalisation and Multiple Representation.* http://aci.ign.fr/Moscow/program.php. Accessed 01.02.2011.

Lüscher, P., & Weibel, R. (2010). Semantics Matters: Cognitively Plausible Delineation of City Centres from Point of Interest Data. *Geographic Information on Demand. 13th Workshop of the ICA commission on Generalisation and Multiple Representation*. http://aci.ign.fr/2010_Zurich/papers.php. Accessed 08.02.2011.

Mackaness, W. A. (2006). Automated Cartography in a Bush of Ghosts. *Cartography and Geographic Information Science, 33*(4), 245–256.

Mackaness, W. A. (2007). Understanding Geographic Space. In W. A. Mackaness, A. Ruas, & T. Sarjakoski (Eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications* (pp. 1–10). Amsterdam: Elsevier.

Mackaness, W. A., & Edwards, G. (2002). The Importance of Modelling Pattern and Structure in Automated Map Generalisation. *Joint ISPRS/ICA workshop on Multi-Scale Representations of Spatial Data.* http://www.ikg.uni-hannover.de/isprs/Program_final.html. Acessed 07.12.2010.

Mallenby, D. (2008). Handling Vagueness in Ontologies of Geographical Information. (Unpublished doctoral dissertation). University of Leeds, Leeds.

Mark D. M. (1989). Conceptual Basis for Geographic Line Generalization. *Auto-Carto 9 Proceedings. Ninth International Symposium on Computer-Assisted Cartography* (pp. 68–77). Falls Church, VA: Amercian Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping.

Mark, D. M. (1993). Toward a Theoretical Framework for Geographical Entity Types. In A. U. Frank & I. Campari (Eds.), *Spatial Information Theory. A Theoretical Basis for GIS, COSIT'93.* Lecture Notes in Computer Science, Vol. 716 (pp. 270–283). Berlin / Heidelberg: Springer-Verlag.

Mark, D. M., & Turk, A. G. (2003). Landscape Categories in Yindjibarndi: Ontology, Environment, and Language. In W. Kuhn, M. Worboys, & S. Timpf (Eds.), *Spatial Information Theory, Foundations of Geographic Information Science, International Conference, COSIT 2003.* Lecture Notes in Computer Science, Vol. 2825 (pp. 28–45). Berlin / Heidelberg: Springer-Verlag.

Mann, P. H. (1965). *An approach to urban sociology.* London: Routeledge.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., & Schneider, S. (2003). *WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies Preliminary Report.* http://wonderweb.semanticweb.org/deliverables.shtml. Accessed 24.01.2011.

Mennis, J. L., Peuquet, D. J., & Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographic Information Science, 14*(6), 501–520.

McMaster, R. B., & Sheppard, E. (2004). Introduction: Scale and Geographic Inquiry. In E. Sheppard, & R. B. McMaster (Eds.), *Scale & Geographic Inquiry. Nature, Society, & Method* (pp. 1–22). Malden, MA: Blackwell Publishing Ltd.

McMaster, R. B., Usery, E. L. (Eds.). (2004). *A Research Agenda for Geographic Information Science.* Boca Raton: CRC Press.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The Psychology of Computer Vision* (pp. 211–277). New York: McGraw-Hill.

Mizen, H., Dolbear, C., & Hart, G. (2005). Ontology Ontogeny: Understanding how an Ontology is created and developed. In M. A. Rodríguez, I. F. Cruz, M. J. Egenhofer, & S. Levashkin (Eds.), *GeoSpatial Semantics. First International Conference, GeoS 2005*. Lectures in Computer Science, Vol. 3799 (pp. 15–29). Berlin / Heidelberg: Springer-Verlag.

Molenaar, M. (2004). Spatial Data Modelling, Dual Partitions and the Specification of Semantics. *Proceedings of the XXth ISPRS Congress*, IAPRS, Vol. XXXV, part B4. http://www.isprs.org/proceedings/XXXV/congress/comm4/comm4.aspx. Accessed 25.01.2011.

Montello, D. R. (1993). Scale and multiple psychologies of space. In A. U. Frank & I. Campari (Eds.), *Spatial Information Theory. A Theoretical Basis for GIS. COSIT'93.* Lecture Notes in Computer Science, Vol. 716 (pp. 312–321). Berlin / Heidelberg: Springer-Verlag.

Montello, D. R. (2003). Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of Geographic Information Science* (pp. 173–189). London: Taylor & Francis.

Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's Downtown?: Behavioral Methods for Determining Referents of Vague Spatial Queries. *Spatial Cognition & Computation, 3*(2), 185–204.

Muller, J.-C. (1991). Generalization of Spatial Databases. In D. J. Maguire, M. Goodchild & D. Rhind (Eds.), *Geographical Information Systems* (pp. 457–475). London: Longman Scientific.

Müller, J. C., Weibel, R., Lagrange, J. P., & Salgé, F. (1995). Generalization: state of the art and issues. In J. C. Müller, J. P. Lagrange, & R. Weibel (Eds.), *GIS and Generalization. Methodology and Practice* (pp. 3–17). London: Taylor & Francis.

Murray, K., Munday, B., & Bush, I. (2005). Enabling Information Integrity within Spatial Data Infrastructures - The Digital National Framework Concept. *FIG Working Week 2005 and GSDI-8*. http://www.gsdi.org/gsdiconf/gsdiconfproceedings/gsdi-8/gsdi-8. Accessed 04.09.2011.

Mustière, S., & Moulin, B. (2002). What is Spatial Context in Cartographic Generalisation? *Proceedings of the Symposium on Geospatial Theory, Processing and Applications.* http://www.isprs.org/proceedings/XXXIV/part4/paper.html. Accessed 30.03.2011.

Mustière, S., Saitta, L., & Zucker, J.-D. (2000). Abstraction in Cartographic Generalization. In Z. W. Ras & S. Ohsuga (Eds.), *Foundations of Intelligent Systems. 12th International Symposium, ISMIS 2000*. Lecture Notes in Artificial Intelligence, Vol. 1932 (pp. 638–644). Berlin / Heidelberg: Springer-Verlag.

Neun, M. (2007). Data Enrichment for Adaptive Map Generalization Using Web Services. (Unpublished doctoral dissertation). University of Zurich, Zurich.

Neun, M., Burghardt, D., & Weibel, R. (2008). Web service approaches for providing enriched data structures to generalisation operators. *International Journal of Geographical Information Science, 22*(2), 133–165.

Nyerges, T. (1991). Representing geographical meaning. In B. P. Buttenfield & R. B. McMaster (Eds.), *Map Generalization: Making Rules for Knowledge Representation* (pp. 59–85). Essex: Longman Scientific and Technical.

Ordnance Survey (2004). *The Digital National Framework – A White Paper*. http://www.dnf.org/. Accessed 04.09.2011.

Ormsby, D., & Mackaness, W. A. (1999). The Development of Phenomenological Generalization Within an Object-oriented Paradigm. *Cartography and Geographic Information Science, 26*(1), 70–80.

O'Sullivan, L., Bovet, S., & Streilein, A. (2008). TLM – The Swiss 3D Topographic Landscape Model. *Proceedings of the XXIst ISPRS Congress*, IAPRS, Vol. XXXVII, part B4 (pp. 1715–1720). http://www.isprs.org/proceedings/XXXVII/congress/tc4.aspx. Accessed 11.03.2011.

Pacione, M. (2005). *Urban Geography* (2nd ed.). Abingdon, UK: Routeledge.

Parker, C. (2004). Research Challenges for a Geo-Information Business. *The Cartographic Journal, 41*(2), 131–141.

Penn, A. (2003). Space Syntax and Spatial Cognition. Or Why the Axial Line? *Environment and Behavior, 35*(1), 30–65.

Penn, A., Hillier, B., Banister, D., & Xu, J. (1998). Configurational modeling of urban movement networks. *Environment and Planning B, 25*(1), 59–84.

Petzold, I., Burghardt, D., & Bobzien, M. (2006). Workflow Management and Generalisation Services. *9th Workshop of the ICA Commission on Generalisation and Multiple Representation.* http://aci.ign.fr/Portland/program.php. Accessed 01.12.2010.

Pointer, G. (2005). The UK's major urban areas. In Office for National Statistics (ONS) (Ed.), *Focus on people and migration: 2005* (pp. 45–60). http://www.statistics.gov.uk/downloads/theme_compendia/fom2005/03_fopm_urbanareas.pdf

Pu, S. (2008). Automatic building modeling from terrestrial laser scanning. In P. van Oosterom, S. Zlatanova, F. Penninga, & E. M. Fendel (Eds.), *Advances in 3D Geoinformation Systems* (pp. 147–160). Berlin / Heidelberg: Springer-Verlag.

Purves, R. S., Clough, P., Jones, C. B., Arampatzis, A., Bucheri, B., Finch, D., … Yang, B. (2007). The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International Journal of Geographical Information Science, 21*(7), 717–745.

Raubal, M. (2005). Mappings For Cognitive Semantic Interoperability. In F. Toppen & M. Painho (Eds.), *Proceedings of the 8th Conference on Geographic Information Science (AGILE '05)* (pp. 291–296). Lisboa, Portugal: Instituto Geográfico Português (IGP).

Regnauld, N. (1996). Recognition of Building Clusters for Generalization. In M. J. Kraak & M. Molenaar (Eds.), *Advances in GIS Research II: Proceedings of the Seventh International Symposium on Spatial Data Handling* (pp 185–198). London: Taylor & Francis.

Regnauld, N. (2001). Contextual Building Typification in Automated Map Generalization. *Algorithmica, 30*(2), 312–333.

Regnauld, N. (2007). A Distributed System Architecture to Provide On-Demand Mapping. *Proceedings of the XXIIIth International Cartographic Conference (ICC 2007).* http://icaci.org/documents/ICC_proceedings/ICC2007/html/Proceedings.htm. Accessed 15.03.2011.

Regnauld, N., & Revell, P. (2007). Automatic Amalgamation of Buildings for Producing Ordnance Survey® 1:50 000 Scale Maps. *The Cartographic Journal 44*(3), 239–250.

Robinson, V. B. (2003). A Perspective on the Fundamentals of Fuzzy Sets and their Use in Geographic Information Systems. *Transactions in GIS, 7*(1), 3–30.

Rodríguez, M. A., & Egenhofer, M. J. (2004). Comparing geospatial entity classes: an asymmetric and contextdependent similarity measure. *International Journal of Geographic Information Science, 18*(3), 229–256.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Ruas, A., & Plazanet, C. (1996). Strategies for automated generalization. In M. J. Kraak and M. Mollenaar (Eds.), *Advances in GIS Research II, Proceedings 7th International Symposium on Spatial Data Handling* (pp. 319–336). London: Taylor & Francis.

Santos, P., Bennett, P., & Sakellariou, G. (2005). Supervaluation Semantics for an Inland Water Feature Ontology. In L. P. Kaelbling & A. Saffiotti (Eds.), *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI 2005)* (pp. 564–569). Professional Book Center.

Scheider, S., Janowicz, K., & Kuhn, W. (2009). Grounding Geographic Categories in the Meaningful Environment. In K. S. Hornsby, C. Claramunt, M. Denis, & G. Ligozat (Eds.), *Spatial Information Theory, 9th International Conference, COSIT 2009.* Lecture Notes in Computer Science, Vol. 5756 (pp. 69–87). Berlin / Heidelberg: Springer-Verlag.

Schuler, M., Dessemontet, P., & Joye, D. (2005). *Die Raumgliederungen der Schweiz.* Neuchâtel: Bundesamt für Statistik (BFS).

Schuurman, N. (2006). Formalization Matters: Critical GIS and Ontology Research. *Annals of the Association of American Geographers, 96*(4), 726–739.

Sen, S. (2007). Two Types of Hierarchies in Geospatial Ontologies. In F. Fonseca, M. A. Rodríguez, & S. Levashkin (Eds.), *GeoSpatial Semantics, Second International Conference, GeoS 2007*. Lecture Notes in Computer Science, Vol. 4853 (pp. 1–19). Berlin / Heidelberg: Springer-Verlag.

Sester, M. (2000). Knowledge acquisition for the automatic interpretation of spatial data. *International Journal of Geographical Information Science, 14*(1), 1–24.

Simpson, J., & Weiner, E. (1989). *The Oxford English Dictionary.* Oxford: Clarendon Press.

Smith, B. (1998). An Introduction to Ontology. In D. Peuquet, B. Smith, & B. Brogaard, (Eds.), *The Ontology of Fields* (pp. 10–14). Bar Harbor, ME: NCGIA.

Smith, B., & Mark, D. M. (1998). Ontology and Geographic Kinds. In T. K. Poiker & N. Chrisman (Eds.), *Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98)* (pp. 308–320). Vancouver: International Geographical Union.

Smith, B., & Mark, D. M. (2001). Geographical categories: an ontological investigation. *International Journal of Geographical Information Science, 15*(7), 591–612

Smith, B., & Mark, D. M. (2003). Do mountains exist? Towards an ontology of landforms. *Environment and Planning B, 30*(3), 411–427.

Sowa, J. F. (2008). Conceptual Graphs. In F. van Harmelen, V. Lifschitz, & B. Porter (Eds.), *Handbook of Knowledge Representation* (pp. 213–237), Amsterdam: Elsevier.

Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA, USA: Brooks Cole Publishing Co.

Sowa, J. F. (1992). *Semantic Networks.* In S. C. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence*. New York, NY, USA: John Wiley & Sons, Inc.

Steiniger, S. (2007). Enabling Pattern-Aware Automated Map Generalization. (Unpublished doctoral dissertation). University of Zurich, Zurich.

Steiniger, S., Lange, T., Burghardt, D. & Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS, 12*(1), 31–59.

Steiniger, S., Taillandier, P. & Weibel, R. (2010). Utilising urban context recognition and machine learning to improve the generalisation of buildings. *International Journal of Geographical Information Science, 24*(2), 252–283.

Steiniger, S., & Weibel, R. (2007). Relations among Map Objects in Cartographic Generalization. *Cartography and Geographic Information Science, 34*(3), 175–197.

Stoter, J., & Salzmann, M. (2003). Towards a 3D cadastre: where do cadastral needs and technical possibilities meet? *Computers, Environment and Urban Systems 27*(4), 395–410.

Sui, D. Z. (1992). A fuzzy GIS modeling approach for Urban land evaluation. *Computers, Environment and Urban Systems*, 16(2), 101–115.

Tallon, A. R., & Bromley, R. D. F. (2004). Exploring the attractions of city centre living: evidence and policy implications in British cities. *Geoforum, 35*(6), 771–787.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2004). Understanding mobile contexts. *Personal Ubiquitous Computing, 8*(2), 135–143.

Teller, J. (2007). Ontologies for an Improved Communication in Urban Development Projects. In J. Teller , J. R. Lee, C. Roussey (Eds.), *Ontologies for Urban Development.* Studies in Computational Intelligence, Vol. 61 (pp. 1–14). Berlin / Heidelberg: Springer-Verlag.

Third, A., Bennett, B., & Mallenby, D. (2007). Architecture for a Grounded Ontology of Geographic Information. In F. Fonseca, M. A. Rodríguez, & S. Levashkin (Eds.), *GeoSpatial Semantics, Second International Conference, GeoS 2007*. Lecture Notes in Computer Science, Vol. 4853 (pp. 36–50). Berlin / Heidelberg: Springer-Verlag.

Thomson, M. K. (2009). Dwelling *on* Ontology – Semantic Reasoning over Topographic Maps. (Unpublished doctoral dissertation). University College London, London.

Thomson, M. K., & Béra, R. (2007). Relating Land Use to the Landscape Character: Toward an Ontological Inference Tool. In A. C. Winstanley (Ed.), *GISRUK 2007. Proceedings of the Geographical Information Science Research UK Conference* (pp. 83–87). Maynooth, Ireland: National University of Ireland Maynooth.

Thomson, M. K., & Béra, R. (2008). A Methodology for Inferring Higher Level Semantic Information from Spatial Databases. In D. Lambrick (Ed.), *Proceedings of the GIS Research UK 16th Annual conference. GISRUK 2008* (pp. 268–274). Manchester: Manchester Metropolitan University.

Touya, G. (2010). Relevant Space Partitioning for Collaborative Generalisation. *Geographic Information on Demand. 13th Workshop of the ICA commission on Generalisation and Multiple Representation*. http://aci.ign.fr/2010_Zurich/papers.php. Accessed 14.03.2011.

Tuan, Y.-F. (1975). Place: An Experiential Perspective. *The Geographical Review, 65*(2), pp. 151–165.

Urban Design Compendium. (2011). http://www.urbandesigncompendium.co.uk/. Acessed 10.01.2011.

Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *The Knowledge Engineering Review, 11*(2), 93–136.

Van Smaalen, J. W. N. (1996). Spatial Abstraction Based on Hierarchical Re-classification. *Cartographica, 33*(1), 65–73.

Van Smaalen, J. W. N. (2003). Automated Aggregation of Geographic Objects. A New Approch to the Conceptual Generalisation of Geographic Databases. (Unpublished doctoral dissertation). Wageningen University, Wageningen.

Varzi, A. C. (2001). Vagueness in Geography. *Philosophy & Geography, 4*(1), 49–65.

Vögele, T., Schlieder, C. & Visser, U. (2003). Intuitive Modelling of Place Name Regions for Spatial Information Retrieval. In W. Kuhn, M. Worboys, & S. Timpf (Eds.), *Spatial Information Theory, Foundations of Geographic Information Science, International Conference, COSIT 2003.* Lecture Notes in Computer Science, Vol. 2825 (pp. 239–252). Berlin / Heidelberg: Springer-Verlag.

Weibel, R. (1997). Generalization of Spatial Data – Principles and Selected Algorithms. In M. van Kreveld, J. Nievergelt, T. Roos, & P. Widmayer (Eds.), *Algorithmic Foundations of Geographic Information Systems* (pp. 99–152). Berlin / Heidelberg: Springer-Verlag.

Weibel, R., & Dutton, G. (1999). Generalising spatial data and dealing with multiple representations. In P. A. Longly, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical Information Systems. Volume 1, Principles and Technical Issues* (2nd ed.) (pp. 125–155). New York: John Wiley & Sons.

Weibel, R., Keller, S., & Reichenbacher, T. (1995). Overcoming the Knowledge Acquisition Bottleneck in Map Generalization: The Role of Interactive Systems and Computational Intelligence. In A. U. Frank & W. Kuhn (Eds.), *Spatial Information Theory. A Theoretical Basis for GIS. Proceedings International Conference COSIT '95* (pp. 139–156). Berlin: Springer.

White, R., & Engelen, G. (2000). High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems, 24*(5), 383–400.

Whitehand, J. W. R. (1992). Recent Advances in Urban Morphology. *Urban Studies, 29*(3), 619–636.

Whitehand, J. W. R., & Whitehand, S. M. (1984). The physical fabric of town centres: the agents of change. *Transactions of the Institute of British Geographers, 9*(2), 231–247.

Williamson, T. (1994). *Vagueness*. London: Routeledge.

Winter, S. (2001). Ontology: buzzword or paradigm shift in GI science? *International Journal of Geographical Information Science, 15*(7), 587–590.

Winter, S., Kuhn, W., & Krüger, A. (2009). Guest Editorial: Does Place Have a Place in Geographic Information Science? *Spatial Cognition & Computation, 9*(3), 171–173.

Yang, B., Luan, X., & Li, Q. (2010). An adaptive method for identifying the spatial patterns in road networks. *Computers, Environment and Urban Systems, 34*(1), 40–48.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control, 8*, 338–353.

Part II

Research Papers

Publication I

Lüscher, P., Burghardt, D., & Weibel, R. (2007). Ontology-driven Enrichment of Spatial Databases. *10th ICA Workshop on Generalisation and Multiple Representation*, Moscow, Russia, August 2–3, 2007.

Ontology-driven Enrichment of Spatial Databases

Patrick Lüscher, Dirk Burghardt, and Robert Weibel
Department of Geography, University of Zurich, CH-8057 Zurich, Switzerland
E-mail: {luescher, burg, weibel}@geo.unizh.ch

Keywords: Spatial data enrichment, pattern recognition, ontologies, urban structures

Abstract

Generalization is an abstraction process by which characteristics of spatial patterns should be preserved and highlighted. This requires the patterns to be detected beforehand. Additionally, automated enrichment of spatial data is of growing importance for many mapping agencies in order to respond to varying user needs. In this paper we present a framework for pattern recognition in urban environments that complements current algorithm-centered approaches by first formalizing spatial patterns in ontologies, and then deductively triggering appropriate low-level pattern recognition techniques. We start our paper by giving an introduction to the terminology of ontologies. Existing work on pattern recognition using semantic models is reviewed. We then outline our general framework and exemplify an ontological model of an urban structure for a case study we are currently working on. Finally, we discuss issues, benefits and challenges of the approach.

1. Introduction

Patterns play an important role during the generalization process: Since their characteristics need to be preserved, they provide a basis for an appropriate selection and parameterization of generalization algorithms. However, most of the spatial databases that exist today have been designed to serve multiple purposes and hence concentrate on the ‘least common denominator’. Data models are usually simple in the sense that they define basic features such as buildings and roads. Therefore, existing databases have to be enriched with patterns that have to be extracted by means of automated pattern recognition techniques (Brassel & Weibel 1988; Ruas & Plazanet 1996).

For mapping agencies, automated enrichment of existing spatial databases with specific higher level concepts allows responding better to customer needs and is therefore useful for many applications. Some concrete examples for the urban domain might be the derivation of the construction period of particular buildings to infer the typical copper concentration per building, a more advanced application might be to connect patterns with urban evolution processes (Camacho-Hübner & Golay 2007), or improved adaptation in mobile services such as navigation by considering spatial contexts specified in the database (Winter 2002).

In the urban context, many specialized pattern recognition algorithms have been employed for detection of structures (Regnauld 1996; Barnsley & Barr 1997; Anders et al. 1999; Boffet 2001; Heinzle et al. 2005; Steiniger 2006a). In the main, these are ‘bottom-up’ in the sense that they first specify a (often visual) pattern to recognise, derive its (geometrical) properties, and use some elaborated detection algorithm (Figure 1, left branch).

Then again, it has been argued that for better adaptation to varying applications, approaches that model the concepts to be derived are needed. For example, it has been pointed out by Mackaness (2006) that abstraction of large-scale databases to very general concepts requires the roles of the individual features and patterns they form to be understood and modeled explicitly. Dutton & Edwardes (2006), Kulik (2005) and Redbrake & Raubal (2004) show the importance of semantic modeling of geographic features in maps to guide user adaptation during generalization.

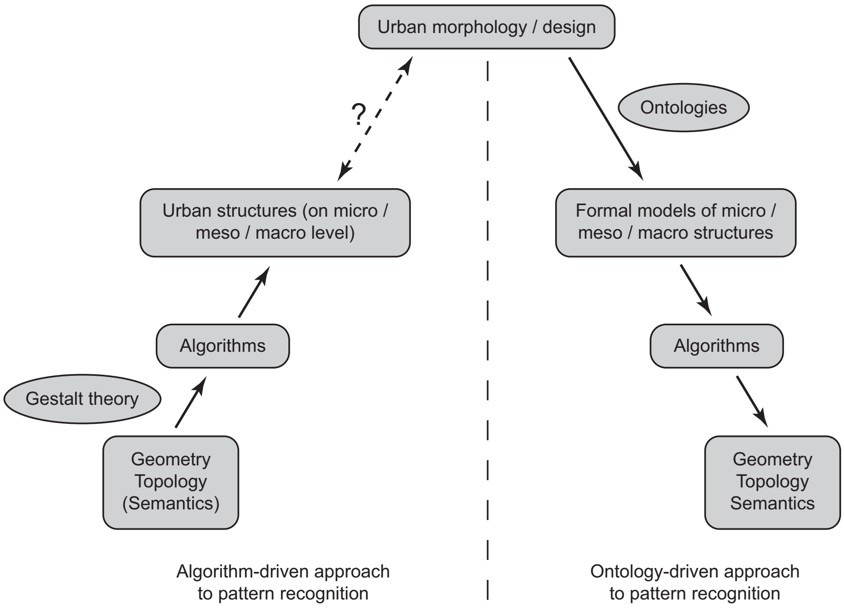


Figure 1. Bottom-up vs. top-down approaches to pattern recognition in urban areas

In our research project we aim at developing methods for the integration of rich semantic concepts into existing spatial databases of the urban domain. The approach we pursue is 'top-down' as shown in Figure 1, right branch: We study the literature on urban morphology and urban design in order to identify specific urban patterns. The next step is to formalize these patterns, their context and hierarchical composition using ontologies. The formal definitions of patterns are then used to deductively trigger appropriate 'low-level' pattern recognition techniques in order to detect them in real databases. We hope that this way we can overcome some important drawbacks of the methods employed nowadays:

Firstly, current pattern recognition methods have often been developed and parameterised for specific databases. However, urban patterns are highly dependent on the cultural background and topographic conditions. For example, the German national atlas (*Nationalatlas Bundesrepublik Deutschland*, Friedrich et al. 2002) describes specific settlement forms (*Angerdorf*, *Hufendorf*, *Gutsdorf*) that cannot be found in other countries such as the UK, which in turn has its own very specific settlement patterns. Therefore, in an ideal approach a domain expert would model important patterns in a formalized language and then have tools available that convert the models automatically to pattern recognition processes.

Secondly, existing pattern recognition algorithms are often not flexible enough to include additional information, such as topography, which may be important to describe the genesis of certain urban forms. Ontologies are a promising means to achieve this integrative role (see Klien & Lutz 2005 for an application example).

Finally, more explanatory power will be contained in the final classifications, since a natural language description of the model can be generated upon request. The network of interlinked concepts can be used for versatile abstraction processes.

The structure of this paper is as follows: After an introduction to the terminology of ontologies (§ 2), we will give an overview of related research in pattern recognition using ontologies (§ 3). We will then state the methodology of our approach and the research issues connected to it (§ 4). Finally, we draw some conclusions of our preliminary work and report on our current and future work on this topic (§ 5).

2. Ontological Modeling

Since ontologies are used in many different contexts, we want to first clarify our understanding of the term. The roots of ontologies lie in philosophy, where the term Ontology is understood as "the science or study of being". It is a specification of "what constitutes reality" in the form of taxonomies (Agarwal 2005). It is independent of epistemology, and since there can be only one reality, there is also only one Ontology, hence the big 'O' and the singular use of the term.

In the last decade, ontologies have attracted large interest in the artificial intelligence community. In AI, an ontology is understood as an explicit specification of a conceptualization (Gruber 1993). A conceptualization is an abstract, simplified view of the world that we want to represent for some reason. Each concept has a concept name (e.g., *ResidentialHouse*), some properties ('number of floors', 'area'), and a set of relations (Rodríguez & Egenhofer 2004).

While this definition reveals some similarities to classic object-oriented modeling, there are some significant differences: Firstly, ontologies are linked hierarchically to higher-level ontologies such that the semantics of concepts is globally clearly defined (Section 2.1). Secondly, concepts in ontologies are rich in semantically defined relations to other concepts (Section 2.2). Thirdly, ontologies can be specified in machine-interpretable languages that allow automatic inference (Section 2.3). Therefore, while object-oriented models define relations on *data*, ontologies define terms with which to represent *knowledge* (Gruber 1993).

2.1 Levels of ontologies

There exists no universally accepted classification of ontologies. For our purposes, we distinguish between three types according to the specialisation of the represented concepts that is similar to the one defined in Guarino (1998) and Fonseca et al. (2002):

- Top-level ontologies: They define very general concepts such as space, time, matter, object, event, action, etc. which are independent of a specific domain or problem. One example of top-level ontology is the SNAP/SPAN ontology by Grenon & Smith (2004) that generally distinguishes between two types of entities. On the one hand objects have a continuous existence through time. On the other hand processes, events, and activities are bound in time – they exist only in their successive temporal parts or phases (Grenon & Smith 2004).
- Domain ontologies: They describe the terminology of a certain domain (such as medicine), or of a general task. We will describe necessary domain ontologies for urban pattern recognition in Section 4.
- Application ontologies: They describe the terms that are on the one hand dependent on a domain, and on the other hand on a very specific task.

The key point is that every level builds on the terms that have been defined in a higher-level ontology. In our framework, basic terms that are needed to trigger the recognition of higher-level concepts would be described as domain ontology. These basic terms comprise single features such as a residential house, and the necessary spatial relations (connected, adjacent, etc.).

2.2 Types of relations

Thus, an ontology is essentially a set of concepts. Concepts can be associated with each other through relations. When modeling entities with ontologies, we can distinguish three types of relations (Rodríguez & Egenhofer 2004 and Fonseca et al. 2002):

- Taxonomic relations: These define sub-concepts and thus create a hierarchy of concepts. For instance, a single family home is a sub-concept of *ResidentialHouse*, which is again a sub-concept of the general concept *Building*.
- Roles: They allow adapting ontologies to specific user views by dynamically assigning concepts to each other. For example, the role *spatialFootprint* for a *Building* can be either played by a polygon, or by a point.
- Partonomic relations: With partonomic relations, aggregate concepts can be defined from a set of basic concepts. Thus, a R*esidentialNeighbourhood* is composed mainly of instances of the concept *ResidentialBuilding*.

Spatial patterns are aggregate concepts that are characterized by the spatial arrangement of the individual parts. For their description, spatial relations have to be defined additionally. For example, “a floodplain is a meadow that is *adjacent* to a river” (Klien & Lutz 2005). Topological relations like *contains* or *touches* are a special class of spatial relations, but also the statement that several houses *are aligned* can be conceptualized as a spatial relation.

When using ontologies for the classification of real data, one wants to find out whether a specific set of objects satisfies all requirements to be classified as an instance of a specific concept. Hence, spatial relations form predicates that have to be evaluated by mapping them to geospatial processing operations (Peachavanish & Karimi 2007). For example, the topological relations mentioned above can be evaluated by the 9-intersection model (Egenhofer & Herring 1991).

One of the main problems is that spatial relations are often fuzzy and hence, the same semantic relation can have different implementations or parameterisations, depending on the context it is used in. For the above mentioned example of floodplains, *adjacent* actually denotes all areas low enough in order to be flooded by the nearby river. If *adjacent* is implemented as a buffer operation, how large should the buffer width be chosen?

2.3 Reasoning with Description Logics (DL)

Ontologies can be specified in a Description Logics (DL) language. In description logics, generally two types of knowledge are represented (Neumann & Möller 2004): A set of axioms (describing a concept) is referred to as terminological box or as *TBox*; factual (assertional) knowledge about the world is called an *ABox*. Let’s clarify the difference between TBoxes and ABoxes with two examples:

- The definition of a floodplain as “a meadow that is *adjacent* to a river” can be formalized in a DL language and states a concept of the TBox. We can tag all areas in a spatial database that satisfy the definition with “Floodplain”. Hence, these areas are part of the ABox.

- "A football stadium is a sports facility which is used for playing football" (Rodríguez & Egenhofer 2004) defines football stadium as a sub-concept of sports facilities in a TBox. The ABox of a London database comprises Highbury Stadium, Matchroom Stadium, Griffin Park, etc.

DL reasoners allow various types of inferences, of which the following might prove to be of importance to our project (from Neumann & Möller 2004):

- whether a concept is subsumed by another concept
- whether an ABox is consistent w.r.t. a TBox;
- whether an individual is an instance of a concept;
- what are the most-specific atomic concepts of which an individual is an instance;
- what are the instances of a concept;
- what are the individuals filling a role for a specified individual;
- what pairs of individuals are related by a specified role; and
- general queries for tuples of individuals mentioned in ABoxes that satisfy certain predicates (so-called conjunctive queries).

Formalizing urban patterns as ontologies reveals some exciting possibilities: As we hope, reasoners can be used to automatically associate instances with concepts; on the other hand, having an ontology-enriched database (enriched manually, or by another system), we can test whether and to which extent it is consistent with our own description.

3. Related work

We will summarize in this section previous and ongoing work that uses explicit semantic models for recognition of spatial patterns.

For computer vision, Neumann & Möller (2004) present an approach to using a DL for high-level scene interpretation. They point out that there has been a gap between low-level vision, which involves techniques for image segmentation and object recognition, and high-level vision, where interpretation tasks may be highly context dependent and knowledge-intensive. They show how specific configurations of objects constrained by temporal and spatial relations such as a table-laying scene for breakfast can be represented by a Description Logic ALCF(D) and sketch a method for using reasoning services as components for the interpretations.

Notable work on semantics-driven interpretation of spatial data has been done in remote sensing for automatic classification of aerial photographs. De Gunst & Vosselmann (1997) present a model-driven approach for the detection of roads using semantic networks. For instance, a two-lane road can be described by three white lines, where the middle line is dashed. Sester (2000) and Anders & Sester (1997) build semantic models for the automatic interpretation of large-scale databases, i.e. they extract different types of houses, streets, parcels and built-up areas from polygon data. The inductive machine learning algorithm ID3 is used to discover relevant spatial properties and relations in manually tagged data. An approach for combining DL with spatial reasoning to formalize spatial arrangements is presented by Haarslev et al. (1994). They propose to combine the reasoning mechanism with a spatial index in order to speed up calculations.

Many spatial concepts are inherently vague. Santos et al. (2005) use supervaluation semantics to integrate vagueness into logical reasoning. They show a prototype implementation which classifies water bodies according to an inland water feature ontology. The inference process is carried out in Prolog.

Ontologies are a means to achieve semantic interoperability in a distributed environment. In this context, Klien & Lutz (2005) discuss the automatic annotation of existing datasets with concepts defined in an ontology. Their approach emphasises spatial relations between features rather than individual feature properties.

Tina Thomson's work aims at building land use maps from OS MasterMap data. Therefore, she intends to use ontologies to model land use categories according to the specific spatial configurations, compositions, relations and other special characteristics (Thomson 2006).

A project of the Ordnance Survey aimed at identifying fields such as farming land or pasture in OS MasterMap data. They used ontologies in order to describe relevant field properties (Kovacs & Zhou 2007).

4. Ontology-driven pattern recognition

4.1 General approach

In this section we will outline our methodology for investigating the role of ontologies in pattern recognition and the benefit of ontology-enriched spatial databases.

Figure 2 shows the general framework. A domain expert (cartographer or urbanist) models the urban structures he/she wants to recognize. The model includes geometrical and semantic components which are needed for their automatic detection and hierarchical composition of patterns, e.g., the pattern might usually be part of an inner city area, which could be either used to restrict the search area for the pattern given inner city areas, or to gain hints for the detection of inner city areas. The model can also include contextual information such as a geographical region for which the pattern is defined, e.g., specific for UK or Israel, and the functional role it plays in a specific context, such as the connection to an urban development process ontology, and thus allow the abstraction to application specific representations.

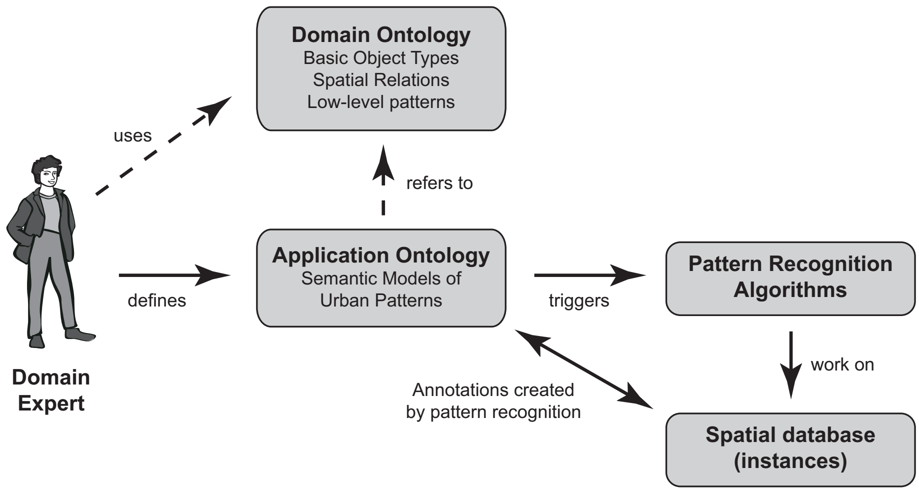


Figure 2. Workflow of the enrichment process using semantic models of urban patterns

These specific models of patterns which we termed 'high-level patterns' constitute application ontologies. We will provide an example for a high-level pattern in the next subsection. In order to be able to define them, a basic vocabulary is needed which is provided as a set of domain ontologies. The 'GIS/cartography' ontology provides concepts for space representation (point, polygon, etc.) and spatial relations (adjacent, within, etc.). There exists also a set of 'low-level patterns' such as alignments and ring structures (buildings), grid patterns and star-shapes (roads), or southern slopes (topography) that are adopted when describing high-level patterns. Another domain ontology is therefore constituted by these low-level patterns.

Ontologies describe a set of concepts and relations between concepts. In order to do the actual data enrichment, a pattern recognition system has to interpret the models and transfer them to a series of spatial processing operations that can be carried out in a GIS environment. To this end, we directly link low-level concepts to spatial algorithms: The pattern recognition system knows how to handle concepts that describe spatial predicates and properties for spatial measures; furthermore, the low-level patterns mentioned above are identified using traditional pattern recognition algorithms. High-level patterns should then be detected automatically by triggering appropriate procedures for measurement of geometrical properties and detection of low-level patterns. Finally, the existing spatial database is annotated with detected low-level and high-level patterns, i.e. links between database objects and concepts are created.

4.2 Formalizing perimeter block developments

In a case study, we are currently working on the formalization of the high-level pattern 'perimeter block developments'. They were a dominant architectural style in Europe from 1880 to 1920 and, as the name implies, perimeter block developments are constituted by buildings that are aligned at the frontage around a rectangular courtyard. Some of the courtyards were originally occupied by workshops, but they were often removed later. Figure 3 shows an extract of a typical perimeter block development area in the City of Zurich.

Figure 3. Typical perimeter block development in the City of Zurich. Source: General plan of Zurich 1:2500.

Figure 4 and 5 show extracts of an ontology that might be built for the urban concept PerimeterBlockDevelopment. We can see that the GIS/Cartography domain ontology also specifies a concept 'Scale', which is important because characteristics of urban structures may depend largely on the scale for which they are defined. For GIS processing functions, it has been proposed that the OGC Simple Feature Specification could be used as a basic domain ontology (Peachavanish & Karimi 2007). The urban morphology defines basic concepts such as urban block or inner city area, which are defined as sub-concepts of Micro- and MesoStructures, respectively. The arrows denote semantic relations of the concept PerimeterBlockDevelopment to its geographical and architectural context. This may be used for example to extract all areas that are instances of inner city concepts in Europe. Thus, through these links, abstraction processes can be formally defined.

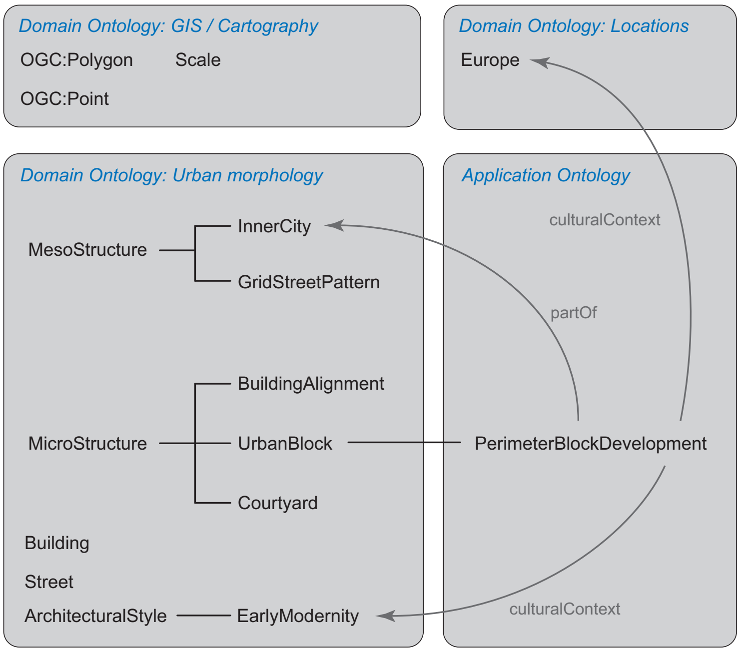


Figure 4. Connection of the concept PerimeterBlockDevelopment with its cultural context

Contextual links allow to flexibly abstract and browse spatial information contained in the database. In order to actually enrich databases with defined concepts, their spatial and functional characteristics have to be encoded in the ontology. Spatial characteristics may include the compositional structures that may be formed from low-level patterns, as well as geometric measures such as typical building sizes. Figure 5 shows a preliminary attempt at linking PerimeterBlockDevelopment to lower-level patterns. Since perimeter block developments typically constitute a grid street pattern, there exists a containment relationship between these concepts. Furthermore, perimeter block developments consist of building alignments, which is also formalized as a containment relationship. A topological relationship between building alignment and street states that the alignments have to be arranged along streets.

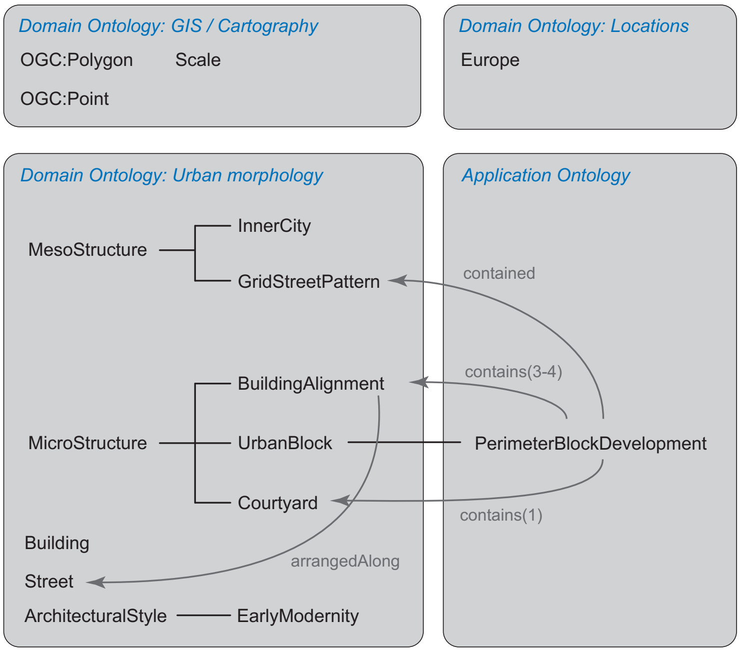


Figure 5. Attempt at linking PerimeterBlockDevelopment to its spatial characteristics

4.3 Research issues

During the first part of our project, the emphasis is on identification and formalization of specific urban concepts. Later, we will have to look at issues concerning the design of the pattern recognition system. Generally, we pursue the following objectives:

1) *Identification and formalization of relevant urban concepts and their spatial properties*. This issue has mainly been addressed by a review of the relevant literature about urban forms and architecture. The formalization of the pattern knowledge is carried out using Protégé (Protégé 2007).

2) *Transformation from ontologies to algorithms that allow their automatic detection in existing spatial databases*. As stated before, we investigate the deployment of automatic reasoning techniques for triggering low-level recognition procedures from ontological descriptions. Commercial reasoners are available off-the-shelf, but they possess no spatial processing capabilities. Reasoners allow to import external functionalities as predicates and functions, so that they can be connected to a GIS environment such as JUMP/JTS (Vivid Solutions 2007).

3) *Actual enrichment of databases with the previously established ontological concepts*. This includes finding an appropriate data model for the connections between ontological concepts and the set of data base objects which instantiate the concepts. Since the concepts (the TBox model) are to be permanently connected to real data (the ABox) which naturally reside in a spatial database, data models have to be found which allow efficient traversal and machine interpretation. It may also be advantageous to store the classification history: If an object is changed during an update, it may affect the patterns it is related to (Haarslev et al. 1994). Another motivation might be that users can retrieve not only patterns, but also the reasons why a concept has been instantiated as such (for example as a textual explanation).

4) *Design of intuitive human-computer interaction methods with the pattern recognition system*: Protégé may be too complex for domain experts. Therefore, we investigate a specific user interface for creating spatial patterns and verify results of detected instances.

4.4 Benefits and challenges of the approach

Compared to the conventional method of building specific algorithms for pattern recognition, our approach has several benefits:

- Properties of patterns are explicitly stated instead of hidden in algorithms. Hence, we will have more explanatory power in the final classifications.
- Pattern recognition will be adaptable to different cultures or contexts by adapting pattern specifications, without actually having to alter the recognition engine.
- Knowledge discovery, representation, and exploitation are integrated within one global framework.
- As already mentioned in Section 2.3, different ways of utilizing the system can be envisioned: On the one hand, it can be used to verify whether a concept is formalized consistently with regard to a certain reality. On the other hand, machine learning techniques can be used for exploring spatial relations that characterize concepts, and hence help domain experts to formalize patterns.

On the other hand, we can identify some issues that may cause difficulties or imply significant drawbacks:

- The semantics of natural language terms denoting spatial relations has been addressed within qualitative spatial reasoning research (Frank 1996). The same term may have different meanings within different contexts (ambiguity of terms), and they are often inherently vague. There is still a lack of knowledge regarding the roles of spatial relations terms in cognitive science research, which may hinder the translation of natural language descriptions into processing chains.
- Similarly, there is also ambiguity and vagueness of concepts. While formalisms to represent ambiguity in ontologies do exist, vagueness has not been profoundly treated so far. The method proposed in Santos et al. (2005) is simplistic since it relies on fixed thresholds. A more natural way to deal with vagueness would be to determine a value of certainty to which a set of objects is trusted to constitute a concept.
- Compared to conventional algorithms, the efficiency of the (spatial) reasoning process may be poor and hence prove to be a significant bottleneck.
- Klien & Lutz (2005) mention that it may not possible to find a fully automated process. In this respect, it is sensible to build a user interface that guides the domain expert through the recognition process and asks for help, where no automatic recognition is possible.

5. Conclusions

In this paper, we investigated the application of ontologies for describing spatial patterns. We believe this would be a sound basis for reasoning about which features and relations are important and hence have to be preserved in automated generalization. In this respect, ontologies are a means to make spatial databases more intelligent. Therefore, methods are needed to connect real data with ontological concepts.

In Section 2, we have introduced the terminology and presented three different levels of ontologies. One conclusion is that application ontologies can be utilized to formalize urban structures.

Section 3 comprises a review about relevant research on spatial pattern recognition using semantic models. As it is pointed out, there has been some work on the conceptual level, but the feasibility for complex real-world problems needs to be proven.

In Section 4, we have presented a methodology for semantic enrichment. The approach is to model high-level concepts in an ontology, whereas low-level pattern recognition procedures are automatically triggered.

The next steps in our work will be to complete the pilot study concerning the perimeter block developments, i.e. to enhance the ontological model and to build a processing chain for their actual detection in spatial databases. Furthermore, we also intend to build a taxonomy of salient urban patterns.

Acknowledgments

The research reported in this paper is part of the PhD project of the first author. Funding by the Swiss State Secretariat for Education and Research (SER) through COST Action C21 (project ORUS, grant no. C05.0081) is gratefully acknowledged.

References

Agarwal, P. (2005): Ontological considerations in GIScience. In: *International Journal of Geographical Information Science*, **19** (5), 501–536.

Anders, K.-H. and Sester, M. (1997): Methods of Data Base Interpretation – Applied to Model Generalization from Large to Medium Scale. In: Förstner, W. and Plümer, L. (Eds.) (1997): *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps. SMATI 97*, Bonn Bad Godesberg, Germany, May 21–23, 1997, 89–103.

Anders, K.-H., Sester, M. and Fritsch, D. (1999): Analysis of Settlement Structures by Graph-Based Clustering. *SMATI'99 - Semantic Modelling for the Acquisition of Topographic Information from Images and Maps*, Munich, Germany, September 7.

Barnsley, M. J. and Barr, S. L. (1997): Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system. In: *Computers, Environment and Urban Systems*, **21** (3–4), 209–225.

Boffet, A. (2001): *Méthode de création d'informations multi-niveaux pour la généralisation cartographique de l'urbain*. Ph.D. thesis, Université de Marne-la-Vallée.

Brassel, K. E. and Weibel, R. (1988): A review and conceptual framework of automated map generalization. In: *International Journal of Geographical Information Systems*, **2** (3), 229–244.

Camacho-Hübner, E. and Golay, F. (2007): Preliminary insights on continuity and evolution of concepts for the development of an urban morphological process ontology. In: Teller, J., Lee, J. R. and Roussey, C. (Eds.) (2007): *Ontologies for Urban Development*. Studies in Computational Intelligence, Vol. 61, 95–108.

de Gunst, M. and Vosselman, G. (1997): A Semantic Road Model for Aerial Image Interpretation. In: Förstner, W. and Plümer, L. (Eds.) (1997): *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps. SMATI 97*, Bonn Bad Godesberg, Germany, May 21–23, 1997, 107–122.

Dutton, G. and Edwardes, A. (2006): Ontological Modeling of Geographical Relations for Map Generalization. *Proceedings of the 9th ICA Workshop on Generalisation and Multiple Representation*, Portland, USA, June 25th, 2006.

Egenhofer, M. J. and Herring, J. R. (1991): Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases. Technical Report, Department of Surveying Engineering, University of Maine, 28 pages.

Fonseca, F. T. et al. (2002): Using Ontologies for Integrated Geographic Information Systems. In: *Transactions in GIS*, **6** (3), 231–257.

Frank, A. U. (1996): Qualitative spatial reasoning: cardinal directions as an example. In: *International Journal of Geographical Information Systems*, **10** (3), 269–290.

Friedrich, K., Hahn B. and Popp, H. (2002): *Nationalatlas Bundesrepublik Deutschland*. Bd. 5. Dörfer und Städte. Spektrum Akademischer Verlag, Berlin, 150 pages.

Grenon, P. and Smith, B. (2004): SNAP and SPAN: Towards Dynamic Spatial Ontology. In: *Spatial Cognition and Computation*, **4** (1), 69–104.

Gruber, T. R. (1993): A translation approach to portable ontology specifications. In: *Knowledge Acquisition*, **5** (2), 199–220.

Guarino, N. (1998): Formal Ontology and Information Systems. In: Guarino, N. (Ed.) (1998): *Formal Ontology in Information Systems. Proceedings of FOIS'98*, Trento, Italy, June 6–8, 1998, 3–15.

Haarslev, V., Möller, R. and Schröder, C. (1994): Combining Spatial and Terminological Reasoning. In: *KI-94: Advances in Artificial Intelligence: 18th German Annual Conference on Artificial Intelligence*, Saarbrücken, Germany, September 8–23, 1994, Lecture Notes in Artificial Intelligence 861, 142–153.

Heinzle, F. et al. (2005): Graph Based Approaches for Recognition of Patterns and Implicit Information in Road Networks. *Proceedings of the 22nd International Cartographic Conference*, A Coruña, Spain, July 11–16, 2005.

Klien, E. and Lutz, M. (2005): The Role of Spatial Relations in Automating the Semantic Annotation of Geodata. In: Cohn, A. G. and Mark, D. M. (Eds.) (2005): *Spatial Information Theory, International Conference, COSIT 2005*, Ellicottville, NY, USA, September 14–18, 2005, LNCS 3693, 133–148.

Kovacs, K. and Zhou, S. (2007): Key challenges in expressing and utilising geospatial semantics at Ordnance Survey. Presentation held at the *European Geoinformatics Workshop*, Edinburgh, UK, March 7–9, 2007. http://www.nesc.ac.uk/action/esi/download.cfm?index=3411. Accessed 22.03.2007.

Kulik, L. et al. (2005): Ontology-Driven Map Generalization. In: *Journal of Visual Languages and Computing*, **16** (3), 245–267.

Mackaness, W. (2002): The Importance of Modelling Pattern and Structure in Automated Map Generalisation. *Proceedings of the Joint ISPRS/ICA Workshop on Multi-Scale Representations of Spatial Data*, Ottawa, Canada, July 7–8, 2002.

Mackaness, W. (2006): Automated Cartography in a Bush of Ghosts. In: *Cartography and Geographic Information Science*, **33** (4), 245–256.

Marshall, S. (2005): *Streets & patterns.* Spon Press, London and New York, 150 pages.

Neumann, B. and Möller, R. (2004): *On Scene Interpretation with Description Logics*. Technical Report, Universität Hamburg, FBI-B-257/04, 30 pages.

OWL (2007): OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-ref/. Accessed 22.03.2007.

Peachavanish, R. and Karimi, H. A. (2007): Ontological Engineering for Interpreting Geospatial Queries. In: *Transactions in GIS*, **11** (1), 115–130.

Protégé (2007): Protégé website. http://protege.stanford.edu/. Accessed 22.03.2007.

Redbrake, D. and Raubal, M. (2004): Ontology-Driven Wrappers for Navigation Services. *Proceedings of the 7th AGILE Conference on GIScience*, Heraklion, Crete/Greece, April 29 - May 1, 2004.

Regnauld, N. (1996): Recognition of Building Clusters for Generalization. In: *Advances in GIS Research II: Proceedings 7th International Symposium on Spatial Data Handling*, Delft, Netherlands, August 12–16, 1996, 4B.1–4B.14.

Rodríguez, M. A. and Egenhofer, M. J. (2004): Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. In: *International Journal of Geographic Information Science*, **18** (3), 229–256.

Ruas, A., and Plazanet, C. (1996): Strategies for automated generalization. In: *Advances in GIS Research II: Proceedings 7th International Symposium on Spatial Data Handling*, Delft, Netherlands, August 12–16, 1996, 6.1–6.17.

Santos, P. et al. (2005): Supervaluation Semantics for an Inland Water Feature Ontology. *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, Edinburgh, Scotland, July 30 - August 5, 2005.

Sester, M. (2000): Knowledge acquisition for the automatic interpretation of spatial data. In: *International Journal of Geographical Information Science*, **14** (1), 1–24.

Steiniger, S. (2006a): Classifying urban structures for mapping purposes using discriminant analysis. In: Priestnall, G. and Aplin, P. (Eds.) (2006): *Proceedings of GISRUK Conference 2006*, Nottingham, UK, April 5-7, 2006, 107–111.

Steiniger, S. et al. (2006b): Recognition of Island Structures for Map Generalization. In: *Proceedings of ACM-GIS'06*, Arlington, Virginia, November 10-11, 2006, 67–74.

Thomson, T. (2006): *Cartographic Data Analysis and Enhancement. Objective 1: Analysis of Functional Information – Needs & Opportunity Analysis, User Requirements, and Use Case.* Technical Report, Department of Geomatic Engineering, University College London.

Torres, M. et al. (2005): Ontology-Driven Description of Spatial Data for Their Semantic Processing. In: Rodríguez, M. A. et al. (Eds.) (2005): *GeoSpatial Semantics: First International Conference, GeoS 2005*, Mexico City, Mexico, November 29–30, LNCS 3799, 242–249.

Vivid Solutions (2007): JUMP/JTS website. http://www.vividsolutions.com/. Accessed 22.03.2007.

Winter, S. (2002): Ontologisches Modellieren von Routen für mobile Navigationsdienste. In: Kelnhofer, F. and Lechthaler, M. (Eds.) (2002): *TeleKartographie und Location-Based Services*. Schriftenreihe der Studienrichtung Vermessungswesen und Geoinformation, Technische Universität Wien, 111–124.

Publication II

Lüscher, P., Weibel, R., & Mackaness, W. (2008). Where is the Terraced House? On The Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases. In A. Ruas & C. Gold, *Headway in Spatial Data Handling. Proceedings of the 13th International Symposium on Spatial Data Handling* (pp. 449–466). Berlin / Heidelberg: Springer-Verlag.

Where is the Terraced House? On the Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases

Patrick Lüscher*[1], Robert Weibel[1] and William A. Mackaness[2]

[1] Department of Geography, University of Zurich
Winterthurerstrasse 190, 8057 Zurich, Switzerland
* Corresponding author: patrick.luescher@geo.uzh.ch

[2] Institute of Geography, School of GeoSciences, University of Edinburgh
Drummond St, Edinburgh EH8 9XP, Scotland, UK

Abstract

In GIS datasets, it is rare that building objects are richly attributed. Yet having semantic information (such as tenement, terraced, semi-detached) has real practical application (in visualisation and in analysis). It is often the case that we can infer semantic information simply by visual inspection – based on metric and topological properties for example. This paper explores the application of pattern recognition techniques as a way of automatically extracting information from vector databases and attaching this information to the attributes of a building. Our methodology builds upon the idea of an ontology-driven pattern recognition approach. These ideas are explored through the automatic detection of terraced houses (based on Ordnance Survey MasterMap® vector data). The results appear to demonstrate the feasibility of the approach. In conclusion we discuss the benefits and difficulties encountered, suggest ways to deal with these challenges, and propose short and long term directions for future research.

Keywords: cartographic databases, ontologies, ontology-driven pattern recognition, building types, geographical characterisation

1 Introduction

Spatial databases currently in use typically have been originally designed and produced in the 1990s. They are rich in geometry, most often include topological structuring, yet they are usually poor in semantics. Those exceptional databases that are semantically rich are restricted to rather narrow purposes – vehicle navigation being a prominent example, where rich additional information on the logics of traffic flow (e.g. one-way streets, pedestrian zones etc.), average speed and speed limits are coded onto the geometry. However, the majority of GIS applications make use of general purpose topographic databases produced either by national mapping agencies (NMAs) or by private companies (e.g. Tele Atlas, NAVTEQ). These general purpose databases are poor in semantics in particular with regards to the representation of higher order semantic concepts that extend beyond the semantics of individual, discrete objects.

This under-representation of semantics limits the utility of the database. The research community has called for methods to automatically 'enrich' such databases. What is required are methods that make *explicit* the spatial relationships and semantic concepts *implicitly* contained in spatial databases. Probably the first research community to call for 'data enrichment' was the map generalisation community (Ruas and Plazanet 1996; Heinzle and Anders 2007). In map generalisation, the special semantics embedded in spatial relations, hierarchical relations, and spatial patterns and structures are critical to modelling the context in which cartographic decisions are made. The map generalisation process utilises information linked to pattern and structure recognition (Brassel and Weibel 1988; Mackaness and Ruas 2007). For example, the decision as to whether to visualise a building on a map will partially depend on contextual information. If it is small yet isolated in a rural area, then the building may be retained and slightly enlarged; if it is in an urban area, it may be eliminated; and if it happens to be a special type of building such as a hospital, it may be replaced by a special symbol (Steiniger 2007).

Generalisation is not the only area where enriched semantics and hence cartographic pattern recognition are crucial. Building types such as tenements or terraced, semi-detached, and detached houses are rarely coded into existing spatial databases, yet, they would provide important semantic information in many practical applications: They give essential clues to prospective house buyers as to what to expect when reading through real estate advertisements (King 1994); information concerning house type is important in planning when trying to develop the right balance between different residential forms in a particular neighbourhood, in quantity surveying or in the recycling of building materials (Müller 2006; Bergsdal et al. 2007). Additionally, enriched semantics can be used to associate urban patterns with urban evolution processes and urban morphology (Camacho-Hübner and Golay 2007); or they may assist adaptation in pedestrian navigation services by considering spatial contexts specified in the database (Winter 2002).

In this paper, we present a novel approach to cartographic pattern recognition. In addition to the more 'traditional' approaches that directly rely on statistical methods and/or geometric algorithms, our approach utilises ontologies to better inform the pattern recognition process and to 'glue' such algorithms together. The paper begins by explaining why ontology-driven pattern recognition has the potential to overcome some of the limitations of traditional approaches and describes the proposed methodology (§ 2). We demonstrate how this approach affords automatic identification of terraced houses from among urban buildings represented in vector form. After presenting an ontology of terraces (§ 3), we explain how the concepts of this ontology can be transformed into an automatic recognition procedure, and we present results of this procedure using Ordnance Survey MasterMap data (§ 4). The paper goes on to identify the benefits and limitations of this technique and suggests ways of overcoming these limitations (§ 5). The conclusion reflects on future research, short and long-term.

2 Ontology-driven Cartographic Pattern Recognition

2.1 Why ontologies are useful in cartographic pattern recognition

Many specialised pattern recognition algorithms have been developed for the detection of structures and patterns specifically in an urban context (e.g. Regnauld 1996; Barnsley and Barr 1997; Anders et al. 1999; Boffet 2001; Christophe and Ruas 2002; Heinzle and Anders 2007; Steiniger et al. 2008). These techniques focus on rather specific patterns that are linked to particular generalisation operations, for instance where we wish to group buildings or to detect alignments in support of aggregation or typification operations (Regnauld 1996; Christophe and Ruas 2002). As there is often an element of fuzziness involved in pattern definitions, these algorithms are often coupled with statistical methods. It remains doubtful whether such algorithms, or a collection thereof, will be sufficient to extract more general, higher order semantic concepts such that we could comprehensively describe the semantics of the morphology of a city. There has to be something additional that enables broader synoptic description of the city form. It has been pointed out by Mackaness (2006) that abstraction from large-scale databases to highly generalised ones requires that the roles of individual features and patterns be understood and modelled explicitly. Dutton and Edwardes (2006), Kulik (2005) and Redbrake and Raubal (2004) show the importance of semantic modelling of geographic features in maps to guide user adaptation during generalisation.

In our research, therefore, we pursued a ‘top-down’ approach to cartographic pattern recognition of urban structures. The individual steps of this ontology-driven approach are illustrated in Figure 1: Based on textual descriptions of urban spaces extracted from the literature, we identify specific urban patterns (step 1); we then formalise these patterns, their context and hierarchical composition based on ontological descriptions (step 2). The ontological definitions of patterns are then used to deductively trigger appropriate ‘low level’ pattern recognition algorithms (step 3) in order to detect them in spatial databases (step 4).

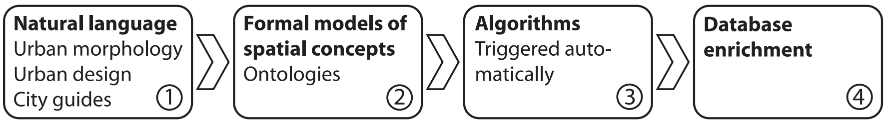


Fig. 1. Steps in the processing chain of ontology-driven pattern recognition

In this way, we can overcome some important drawbacks of methods used today:

- Current pattern recognition methods have often been developed and parameterised for specific data models and databases. For instance, if they have been developed with German ATKIS data in mind, they might assume that roads are represented by centre lines. It is anticipated that ontologies will provide meta-knowledge that improves the ‘interoperability’ and applicability of pattern recognition methods across different databases.
- It is often the case that existing pattern recognition algorithms cannot be adapted to take into account additional information in the detection procedure, such as topography, which may be important in describing the genesis of certain urban patterns. Ontological descriptions help make explicit all the criteria that enable us to identify a particular composition of buildings (Klien and Lutz 2005).
- The nature of geographic form means that many spatial patterns cannot be crisply defined and delineated. Therefore pattern recognition additionally depends upon the use of statistical techniques (e.g. Steiniger et al. 2008). The result of typical statistical methods may be difficult to interpret, however, as the relations that are inferred between pattern variables are purely statistical rather than revealing causes and consequences. Ontologies, on the other hand, represent the concepts that are modelled, as well as the relations between them in an explicit way. Thus, they are inherently more transparent than statistical methods and have potentially more explanatory power.

2.2 Ontologies for cartographic pattern recognition

The term 'ontology' is defined from an engineering science perspective and is defined as an explicit specification of a shared conceptualisation (Gruber 1993). It is thus an attempt to capture the knowledge in a certain domain in a systematic way by breaking it down into the types of entities (*concepts*) that exist and the *relations* that hold between them. Ontologies can be classified according to the degree of formalisation into informal (written in natural language), semi-formal (restricted language), and formal (artificial language) ontologies (Agarwal 2005). An alternate classification is one that conforms to the degree of specialisation and is divided into top-level, domain, and task ontologies, the last being the most specific one (Guarino 1998). While a key application of ontologies is to improve the interoperability between information systems (Fonseca et al. 2002), ontologies are also employed as a method of eliciting knowledge that exists in a domain (Agarwal 2005).

In this research we seek to explain complex urban phenomenon in terms of other, possibly simpler phenomena, such that the meaning of the concept is derived from the meaning of the related concepts. We refer to the first kind as a 'higher order concept', and to the second kind as a 'lower order concept'. The lower order concepts may themselves be composite concepts, in which case they have to be broken down further into still lower order concepts. Alternatively they might be simple in the sense that they can be directly related to cartographic measures or a cartographic structure recognition algorithm.

2.3 Data enrichment using ontologies

The concept above constitutes an *ideal prototype* (a template). Real occurrences of a concept will normally comply only to a certain degree with the template. Hence, a value which expresses the degree of congruence between reality and the ideal prototype of the concept has to be calculated: where $con(C_i, R_j) = 0$ when a realisation R_j differs completely from a template C_i, and $con(C_i, R_j) = 1$ when they match perfectly.

For low order concepts $con(C_i, R_j)$ is extracted by a cartographic pattern recognition algorithm. For composite concepts, which are defined by their relations to lower order concepts, $con(C_i, R_j)$ has to be inferred from the congruence values of their constituting concepts. Here we distinguish between two types of relationships:

- Some relationships, such as the subclass relationship, translate to strict exclusions:

$$con(C_i, R_k) = 0 \rightarrow con(C_j, R_k) = 0 \qquad (1)$$

If C_j is a subclass of C_i. For example, if a spatial object is not a building then it cannot be a terraced house, regardless of the congruence values of the other constituting concepts, since terraced houses are a subclass of buildings.

- For other relationships, congruence values of the constituting values have to be intersected. One possibility for combining single similarity values to an overall value is by calculating a weighted linear average:

$$con(C_i, R_k) = \left(\sum w_j con(C_j, R_k)\right) \Big/ \sum w_j \quad (2)$$

  Where $con(C_j, R_k)$ is the congruence value of a constituent concept of C_i and the weight w_j is an influence value of the subconcept. For reasons of simplicity, all weights were equated to 1 for this study.

Thus, the calculation of congruence values starts with the patterns at the bottom and then propagates iteratively to higher order concepts. This is similar to forward reasoning in description logics. At the end of this process, spatial objects can be annotated with the congruence value for the concepts defined in the ontology.

2.4 Related work

Our review of related work will be brief and will focus exclusively on approaches that use explicit semantic models for the recognition of spatial patterns in *vector databases*, ignoring the literature related to image interpretation and computer vision.

Sester (2000) and Anders and Sester (1997) built semantic models for the automatic interpretation of large-scale vector databases. They extracted different types of houses, streets, parcels and built-up areas from polygon data. The inductive machine learning algorithm ID3 is used to discover relevant spatial properties and relations in manually tagged data. An approach for combining spatial reasoning with description logics to formalise spatial arrangements is presented by Haarslev et al. (1994).

Many spatial concepts are inherently vague. Santos et al. (2005) used supervaluation semantics to integrate vagueness into logical reasoning. They show a prototype implementation in Prolog that classifies water bodies according to an ontology of inland water features.

Ontologies are a means to achieve semantic interoperability in a distributed environment. In this context, Klien and Lutz (2005) discuss the automatic annotation of existing datasets with concepts defined in an ontology. Their approach emphasises spatial relations between features rather than individual feature properties. Thomson (2006) sought to build land use maps from

OS MasterMap data. Her intention was to use ontologies to model land use categories according to the specific spatial configurations, compositions, and relations. This is somewhat similar to a project at the Ordnance Survey which sought to identify fields such as farming land or pasture in OS MasterMap data, using ontologies (Kovacs and Zhou 2007).

We conclude our review with a few observations. First, the amount of work using semantic models for pattern recognition in cartographic vector databases is much smaller than the literature on purely algorithmic approaches. Second, much of the research reviewed in this subsection is restricted to a selected set of spatial patterns; the extensibility and the potential generality of these approaches is rarely discussed. And finally, few references have actually gone into details of instantiating the proposed ontology definitions and of implementing a prototype to prove the validity of the approach; many stay at the more theoretical level.

3 An ontology of terraced houses

«Beyond the mills … were the rows of terraces – mean little houses, with low ceilings and dark cramped rooms.» *— Jane Rogers, Her Living Image.*

In this section we want to show how textual descriptions of urban concepts can be formalised and thus serve as a basis for their detection. The concepts in this study were collected from texts on urban morphology, which is "the study of the physical (or built) fabric of urban form, and the people and processes shaping it" (Jones and Larkham 1991). The hypothesis of urban morphology is that economic and social significance of a town finds its expression in the physiognomy, which is a combination of town plan, pattern of building forms, and pattern of urban land use (Conzen 1969). Concept descriptions were complemented using dictionaries such as the Oxford English Dictionary (Simpson and Weiner 1989). By way of example, Figure 2 shows residential house types identified in the urban morphology literature.

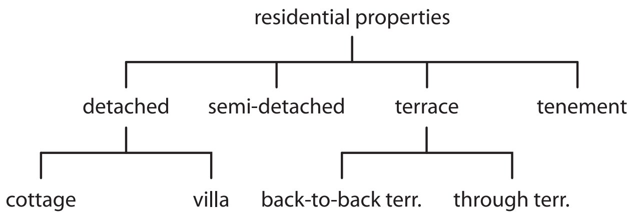


Fig. 2. Urban residential house types extracted from the glossary of urban form (Jones and Larkham 1991)

While 'terraced house' is generally a synonym for 'row house' and may therefore have different features depending on culture and construction period, the *prototype* for our formalisation is the characteristic terrace house settlement in the UK of the late Victorian and Edwardian period. It is linked to the Public Health Act of 1875, established to improve urban living conditions and resulted in re-housing of population from slum clearance areas (Conzen 1969). The demand for cheap mass housing was met by creating rows of unified buildings sharing sidewalls. Because of the low social status of the dwellers, lot sizes and room footprints were small.

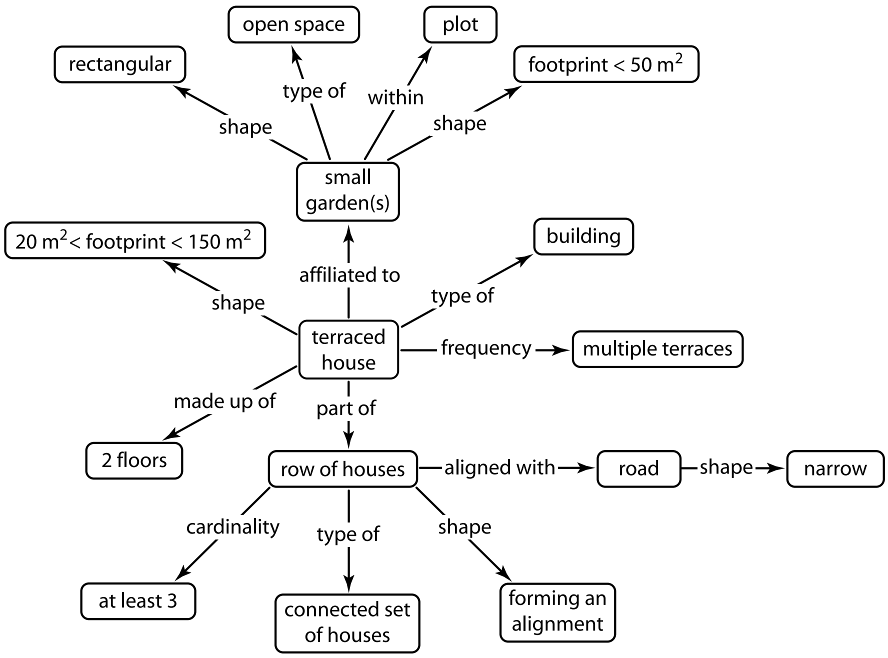


Fig. 3. An ontology of terraced houses

Terraced houses usually have small front-gardens and possibly attached sculleries and a yard at the rear. Often, multiple rows of houses form an area of a highly regular plot pattern. The ontology extracted from these descriptions is shown in Figure 3.

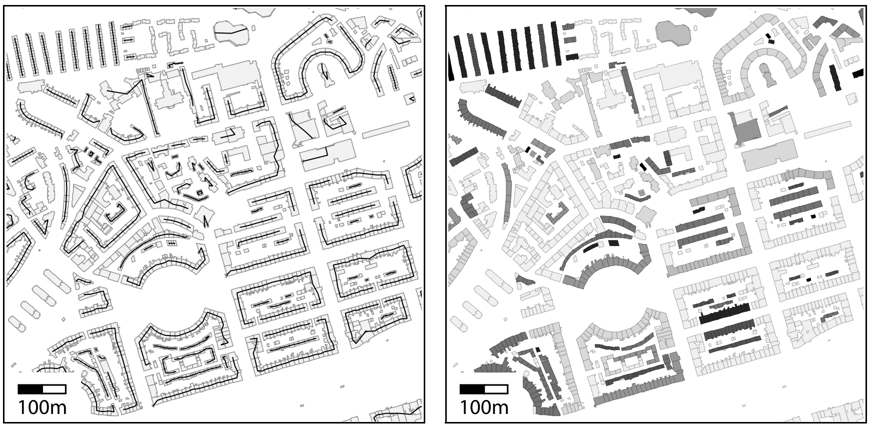
4 Experiment

In order to assess the data enrichment performance of the ontology-driven approach in general and the terraced house ontology in particular an experiment was carried out using OS MasterMap data for Edinburgh, Scotland, UK. OS MasterMap provides a planar topology, that is, space subdivided into polygons such that no polygons overlap, and every location is covered by exactly one polygon. The ontology was realised in a prototype for ontology-driven pattern recognition programmed in Java, tough the current prototype does not yet implement the concepts ‘small garden(s)’ and ‘narrow roads’.

4.1 Extraction and composition of low order concepts

As described in § 2.3, low order concepts can be mapped to cartographic measures. For the terraced house ontology, the following low order concepts have been implemented:

- The concept 'building' can be trivially extracted from OS MasterMap; an attribute encodes whether a polygon represents open land, transportation or a building.
- '$20\ m^2$ < footprint < $150\ m^2$' was obtained using a crisp threshold for building areas.
- Since OS MasterMap does not contain any information on the height of buildings, the concept 'made up of two floors' had to be omitted.
- For the concept 'row of houses', groups of buildings were created. There are several methods that calculate alignments of buildings (see Burghardt and Steiniger 2005 for an overview). We derived the degree of alignment by grouping buildings sharing a common wall and then connecting the centroids of the buildings for groups containing at least three buildings, so that a path representing the general form of the group was formed (Figure 4a). The form of the path was assessed using the compactness of the area covered by the path. We also rated homogeneity of buildings within groups by means of the standard deviation of the building areas. Finally, the form of the path and the homogeneity of buildings were averaged to obtain the congruence value of building groups to alignments. Figure 4b shows the congruence values for an extract of our study area: Linearly arranged, homogeneous blocks in the northwest of the extract achieve high congruence values, whereas 'perimeter-block development'-like blocks receive low congruence values.



(a) (b)

Fig. 4. (a) Paths to qualify the general form of building groups (b) Congruence of buildings to the concept 'row of houses'. Light values denote low, dark values denote high similarity. OS MasterMap data Ordnance Survey © Crown Copyright. All rights reserved.

- The concept ‘multiple terraces’ was derived by identifying the main axes of building groups and clustering these groups using the direction of the axes. The clusters were then qualified by means of the homogeneity of axes directions, length of axes, and homogeneity of buildings within the clusters. To this end, standard deviations were calculated and averaged as previously discussed. Figure 5 shows an example of the clusters found. Note that in the right hand part of the figure, there are two areas – marked (1) and (2) – with regular rows of buildings that have not been classified as ‘multiple terrace’. This is because the footprints of the building areas are too large and hence they correspond rather to tenements than to terraced houses. The two rows marked as (3) have not been detected as being ‘regular’ because we defined that there must be at least three approximately parallel rows of houses for this condition to be met.

Finally, the congruence value of ‘terraced house’ was calculated by intersecting ‘building’, ’20 m^2 < footprint < 150 m^2’, ‘row of houses’, and ‘multiple terraces’ as explained in § 2.3.

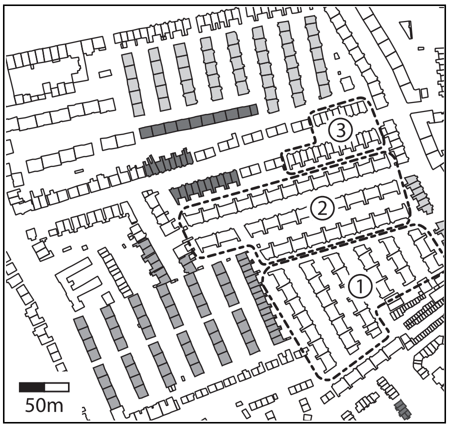


Fig. 5. Areas of multiple terraces. OS MasterMap data Ordnance Survey © Crown Copyright. All rights reserved.

4.2 Results

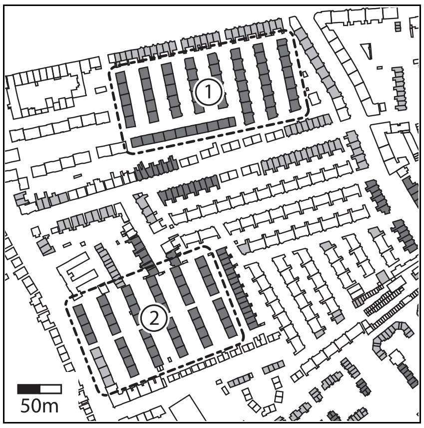
The classification has been carried out for an area covering a part of the City of Edinburgh, 4.6 km x 3.6 km size. The congruence values obtained were deliberately classified into the three categories in order to simplify the validation process:

- ‘high’ congruence: $con(\text{‘terraced house’}, R_i) > 0.8$
- ‘medium’ congruence: $0.6 < con(\text{‘terraced house’}, R_i) \leq 0.75$
- ‘low’ congruence: $con(\text{‘terraced house’}, R_i) \leq 0.6$

Of the 20 990 houses in the study area, 1 557 were classified as having high congruence, 5 064 as having medium congruence, and 14 369 as having low congruence with the concept 'terraced house'. We did some ground truthing to measure the occurrence of terraced houses, but not for all of Edinburgh. The results were compared to ground truth where available, and visually compared to aerial photographs elsewhere.

The algorithm identified six larger areas of terraced houses. Five of those areas correspond to settlements known as the 'Edinburgh Colonies' that fit pretty nicely to our conceptualisation of terraced houses (Figures 6 and 7). There was one settlement of the 'Colonies' that was not classified fully as having a high congruence value, namely the North Forth Street Colony (Figure 7b). The reason for this is that our algorithm for 'multiple terraces' extracts parallel rows of houses rather than orthogonally arranged rows such as in the North Forth Street Colony.

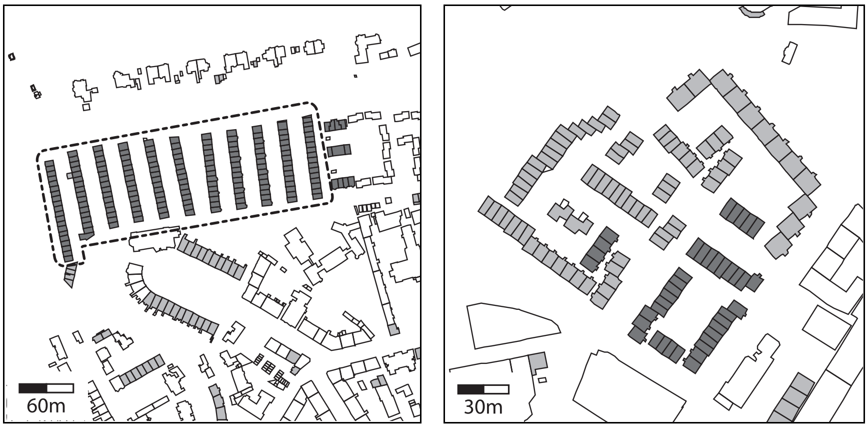
Finally, 775 of the 1 271 buildings classified as having high congruence could be definitively confirmed as terraced houses. This does not imply that the remaining 496 buildings with high congruence values are in fact not terraces (equivalent to an error of commission), but simply that in these cases a ground survey will be needed to confirm the result.



(a)

(b)

Fig. 6. (a) Leith Links (1) and Lochend Road (2) Colonies. (b) Picture of terraced houses in the Leith Links Colony. High congruence with 'terraced house' concept in dark grey, medium congruence in light grey, for low congruence just building boundaries are shown. OS MasterMap data Ordnance Survey © Crown Copyright. All rights reserved.



(a) (b)

Fig. 7. (a) Stockbridge Colony. (b) North Forth Street Colony. Contrast levels as in Figure 6. OS MasterMap data Ordnance Survey © Crown Copyright. All rights reserved.

5 Discussion

5.1 Benefits

In general, the results generated are plausible. This research has shown how textual descriptions of urban patterns can be used to define an ontology that in turn can be used to inform the detection of these patterns, thus enabling enrichment of existing vector cartographic databases. Since the ontology makes the concepts and relations defining a spatial pattern explicit, it can also be used to generate graphical representations such as the one seen in Figure 3 as well as textual descriptions (or metadata) about the extracted patterns. And finally, it follows trivially from Figure 3 that it would be easy to modify concepts in the ontology of the higher order concept 'terraced house', or add further low order concepts to it. For instance, it would be possible to accommodate cultural differences between prototypical terraces in different regions or countries. Our ultimate aim is to extend this framework such that a domain expert can define his/her conceptualisation of *any* urban pattern as an ontology and has a useful set of low order patterns at hand that can be used to perform the detection process.

5.2 Difficulties

Operationalisation of concepts: The operationalisation of lower order patterns is not necessarily easy. One example is the concept 'multiple terraces', which means that a larger number of rows of terraces are arranged regularly. Regularity itself is a loose term, and there are several ways of measuring it. We defined a regular arrangement of terraces as a group of at least three approximately parallel rows of houses. The generation of such groups involves creating a buffer to both sides of each main axis and intersecting this buffer with other main axes. This works well for typical terraced houses (Figure 5), but more general definitions may be needed when different concepts are to be detected.

Another example is the derivation of alignments of houses. There exist various methods for grouping houses into alignments (Burghardt and Steiniger 2005; Christophe and Ruas 2002; Boffet 2001). They assume different conceptualisations of the constitution of alignments and hence produce different results. Therefore, the influence of the choice of implementation of the low order concepts to the inference workflow and to the recognition performance has to be investigated in detail.

Thresholds: Some of the concepts involved setting a threshold (e.g. the area of the footprint of a building). Such crisp thresholds are rather undesirable and could be improved using fuzzy membership functions (Ladner et al. 2003).

Defining a processing order: For complex concepts like terraced houses, a processing hierarchy has to be identified. The hierarchy defines the order of the inference of lower level concepts and their composition into higher level concepts. This is made difficult by the fact that lower level concepts in different sub-branches sometimes depend on each other. For example, the detection of areas of multiple terraces assumes that terraces have already been detected, but in turn also inform the detection process of terraced houses. Since we turned our ontology manually into a detection process, these interdependencies could be accounted for. With respect to a more automated operationalisation process (which is desirable because domain experts are usually not experts in programming), we need more research on how we can formally model such interdependencies.

Alternative ways of concept inference: The method to calculate congruence values of composite concepts was given in § 2.3. The strengths are its simplicity, the fact that the output is a similarity (congruence) value instead of a hard classification, and the high level of transparency of the results. Fuzzy logic would offer a similar but more complex approach.

Supervised classification methods (Steiniger et al. 2008) use training data to define characteristic properties of different classes, and hence there is no need to set thresholds. On the other hand, the performance of supervised classification depends largely on the quality of the training samples used. Furthermore, it is our opinion that using ontologies can better integrate structural knowledge about concepts into the reasoning process and hence is better adapted to detecting complex concepts.

6 Conclusions

In this paper, we have advocated the use of ontologies to better inform the recognition of spatial patterns and structures in the urban environment from cartographic vector databases. We have explained how we envisage ontology-driven cartographic pattern recognition as a novel complement to traditional algorithmic and statistical pattern recognition. For the example of terraced houses, we have developed an ontology, implemented the corresponding recognition procedure in Java, and validated it using OS MasterMap data.

There are several insights that can be gained from this work. Ontologies definitely render the recognition process more flexible (and extensible), enable greater self-documentation, and make us better equipped to compose complex concepts from simple concepts as opposed to traditional algorithmic techniques. Despite the great potential of ontology-driven approaches, they still represent a relatively unfamiliar approach in this application domain and hence pose a series of challenges for future research. Among the difficulties encountered in our study (§ 5) are the operationalisation of concepts; the proper way of dealing with thresholds and fuzziness; dealing with concept interdependencies when integrating simple to complex concepts; and alternative ways of concept inference.

In the short term we plan the following extensions to this study: Complete ground truthing to completely validate our results; application of the procedure to other study areas; modification and/or extension of the ontology of terraced houses (e.g. to accommodate cultural differences); experiments using people to study where and how they visually detect terraces; and development and implementation of ontologies of other house types (semi-detached, detached, tenement). In the mid term we envisage first integrating the different building ontologies to a 'house' ontology, and later to an ontology of even higher order concepts such as 'residential area'. And in the long term we hope to develop methods for the automated 'deployment' of ontologies, which will facilitate the application of ontology-driven pattern recognition for domain experts.

Acknowledgements

Funding of the first author by the Swiss State Secretariat for Education and Research (SER) through COST Action C21 (project ORUS, grant no. C05.0081) is gratefully acknowledged. The OS MasterMap data used in this study were made available to the University of Edinburgh through Digimap. We are also grateful to the Institute of Geography, University of Edinburgh for hosting an extended stay of the Zurich researchers during November 2007.

References

Agarwal P (2005) Ontological Considerations in GIScience. International Journal of Geographical Information Science 19(5):501–536

Anders K-H, Sester M (1997) Methods of Data Base Interpretation – Applied to Model Generalization from Large to Medium Scale. In: Förstner W, Plümer L (eds) Semantic Modeling for the Acquisition of Topographic Information from Images and Maps: SMATI 97. Birkhäuser, Basel, pp 89–103

Anders K-H, Sester M, Fritsch D (1999) Analysis of Settlement Structures by Graph-Based Clustering. SMATI'99 – Semantic Modelling for the Acquisition of Topographic Information from Images and Maps, 7th September, Munich, Germany

Barnsley MJ, Barr SL (1997) Distinguishing Urban Land-use Categories in Fine Spatial Resolution Land-cover Data using a Graph-based, Structural Pattern Recognition System. Computers, Environment and Urban Systems 21(3–4):209–225

Bergsdal H, Brattebø H, Bohne RA, Müller DB (2007) Dynamic Material Flow Analysis of Norway's Dwelling Stock. Building Research & Information 35(5):557–570

Boffet A (2001) Méthode de création d'informations multi-niveaux pour la généralisation cartographique de l'urbain. Ph.D. thesis, Université de Marne-la-Vallée

Brassel KE, Weibel R (1988) A Review and Conceptual Framework of Auto-mated Map Generalization. International Journal of Geographical Information Systems 2(3):229–244

Burghardt D, Steiniger S (2005) Usage of Principal Component Analysis in the Process of Automated Generalisation. Proceedings of the 22nd International Cartographic Conference, 11–16 July, A Coruna, Spain

Camacho-Hübner E, Golay F (2007) Preliminary Insights on Continuity and Evolution of Concepts for the Development of an Urban Morphological Process Ontology. In: Teller J, Lee JR, Roussey C (eds) Ontologies for Urban Development. Studies in Computational Intelligence, Vol. 61. Springer, Berlin Heidelberg New York, pp 95–108

Christophe S, Ruas A (2002) Detecting Building Alignments for Generalisation Purposes. In: Richardson DE, van Oosterom P (eds) Advances in Spatial Data Handling (10th International Symposium on Spatial Data Handling). Springer, Berlin Heidelberg New York, pp 419–432

Conzen MRG (1969) Alnwick, Northumberland: A Study in Town-plan Analysis. Institute of British Geographers, London

Dutton G, Edwardes A (2006) Ontological Modeling of Geographical Relations for Map Generalization. Proceedings of the 9th ICA Workshop on Generalisation and Multiple Representation, 25th June, Portland, USA

Fonseca FT, Egenhofer MJ, Agouris P, Câmara G (2002) Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS 6(3):231–257

Gruber TR (1993) A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5(2):199–220

Guarino N (1998) Formal Ontology and Information Systems. In: Guarino N (ed) Formal Ontology in Information Systems. Proceedings of FOIS'98. IOS Press, Amsterdam, pp 3–15

Haarslev V, Möller R, Schröder C (1994) Combining Spatial and Terminological Reasoning. In: Nebel B, Dreschler-Fischer LS (eds) KI-94: Advances in Artificial Intelligence: 18th German Annual Conference on Artificial Intelligence. Lecture Notes in Artificial Intelligence 861. Springer, Berlin Heidelberg New York, pp 142–153

Heinzle F, Anders K-H (2007) Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks. In: Mackaness WA, Ruas A, Sarjakoski LT (eds) Generalisation of Geographic Information: Cartographic Modelling and Applications. Elsevier Science, Amsterdam et al., pp 233–253

Jones AN, Larkham PJ (1991) Glossary of Urban Form. Historical Geography Research Series no.26. Institute of British Geographers, London

King AD (1994) Terminologies and Types: Making Sense of Some Types of Dwellings and Cities. In: Franck KA, Schneekloth LH (eds) Ordering Space – Types in Architecture and Design. Van Nostrand Reinhold, New York et al., pp 127–144

Klien E, Lutz M (2005) The Role of Spatial Relations in Automating the Semantic Annotation of Geodata. In: Cohn AG, Mark DM (eds) Spatial Information Theory, International Conference, COSIT 2005. Lecture Notes in Computer Science 3693. Springer, Berlin Heidelberg New York, pp 133–148

Kovacs K, Zhou S (2007) Key Challenges in Expressing and Utilising Geospatial Semantics at Ordnance Survey. Presentation held at the European Geoinformatics Workshop, 7–9

March, Edinburgh, UK. http://www.nesc.ac.uk/action/esi/download.cfm?index=3411 Accessed 22.01.2008

Kulik L, Duckham M, Egenhofer MJ (2005) Ontology-Driven Map Generalization. Journal of Visual Languages and Computing 16(3):245–267

Ladner R, Petry FE, Cobb MA (2003) Fuzzy Set Approaches to Spatial Data Mining of Association Rules. Transactions in GIS 7(1):123–138

Mackaness WA (2006) Automated Cartography in a Bush of Ghosts. Cartography and Geographic Information Science 33(4):245–256

Mackaness WA, Ruas A (2007) Evaluation in the Map Generalisation Process. In: Mackaness WA, Ruas A, Sarjakoski LT (eds) Generalisation of Geographic Information: Cartographic Modelling and Applications. Elsevier Science, Amsterdam, pp 89–111

Müller DB (2006) Stock Dynamics for Forecasting Material Flows – Case Study for Housing in The Netherlands. Ecological Economics 59(1):142–156

Redbrake D, Raubal M (2004) Ontology-Driven Wrappers for Navigation Ser-vices. In: Toppen F, Prastacos P (eds) AGILE 2004, 7th Conference on Geographic Information Science. Crete University Press, Heraklion, pp. 195–205

Regnauld N (1996) Recognition of Building Clusters for Generalization. In: Kraak MJ, Molenaar M (eds) Advances in GIS Research II: Proceedings of the Seventh International Symposium on Spatial Data Handling. Taylor & Francis, London, pp 4B.1–4B.14

Ruas A, Plazanet C (1996) Strategies for Automated Generalization. In: Kraak MJ, Molenaar M (eds) Advances in GIS Research II: Proceedings of the Seventh International Symposium on Spatial Data Handling. Taylor & Francis, London, pp 6.1–6.17

Santos P, Bennett B, Sakellariou G (2005) Supervaluation Semantics for an Inland Water Feature Ontology. Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, July 30 – August 5, Edinburgh, Scotland

Sester M (2000) Knowledge Acquisition for the Automatic Interpretation of Spatial Data. International Journal of Geographical Information Science 14(1):1–24

Simpson J, Weiner E (1989) The Oxford English Dictionary. Oxford University Press, Oxford

Steiniger S (2007) Enabling Pattern-aware Automated Map Generalization. Ph.D. thesis, University of Zurich

Steiniger S, Lange T, Burghardt D, Weibel R (2008) An Approach for the Classification of Urban Building Structures Based on Discriminant Analysis Techniques. Transactions in GIS 12(1):31–59

Thomson T (2006) Cartographic Data Analysis and Enhancement. Objective 1: Analysis of Functional Information – Needs & Opportunity Analysis, User Requirements, and Use Case. Technical Report, Dept. of Geomatics Eng., University College London

Winter S (2002) Ontologisches Modellieren von Routen für mobile Navigationsdienste. In: Kelnhofer F, Lechthaler M (eds) TeleKartographie und Location-Based Services. Schriftenreihe der Studienrichtung Vermessungswesen und Geoinformation, Technische Universität Wien, pp 111–124

Publication III

Lüscher, P., Weibel, R., & Burghardt, D. (2009). Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. *Computers, Environment and Urban Systems, 33*(5), 363–374.

Computers, Environment and Urban Systems 33 (2009) 363–374

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data

Patrick Lüscher[a,*], Robert Weibel[a], Dirk Burghardt[b]

[a] Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[b] Institut für Kartographie, Technische Universität Dresden, Helmholtzstraße 10, 01062 Dresden, Germany

ARTICLE INFO

Keywords:
Pattern recognition
Spatial database enrichment
Ontology
Bayesian network
Supervised classification
Building types

ABSTRACT

This paper presents an ontology-driven approach for spatial database enrichment in support of map generalisation. Ontology-driven spatial database enrichment is a promising means to provide better transparency, flexibility and reusability in comparison to purely algorithmic approaches. Geographic concepts manifested in spatial patterns are formalised by means of ontologies that are used to trigger appropriate low level pattern recognition techniques. The paper focuses on inference in the presence of vagueness, which is common in definitions of spatial phenomena, and on the influence of the complexity of spatial measures on classification accuracy. The concept of the English terraced house serves as an example to demonstrate how geographic concepts can be modelled in an ontology for spatial database enrichment. Owing to their good integration into ontologies, and their ability to deal with vague definitions, supervised Bayesian inference is used for inferring complex concepts. The approach is validated in experiments using large vector datasets representing buildings of four different cities. We compare classification results obtained with the proposed approach to results produced by a more traditional ontology approach. The proposed approach performed considerably better in comparison to the traditional ontology approach. Besides clarifying the benefits of using ontologies in spatial database enrichment, our research demonstrates that Bayesian networks are a suitable method to integrate vague knowledge about conceptualisations in cartography and GIScience.

1. Introduction

Spatial databases currently produced by national mapping agencies (NMAs) are typically modelled closely after the original map products which they replaced, meaning that they are rich in geometry but poor in semantics, particularly with regards to the representation of higher order geographic concepts that extend beyond the semantics of individual, discrete objects. Examples of geographic concepts that are not coded in current spatial databases include the geomorphological process underlying stretches of a coastline (estuary, fjord, skerry etc.), the extent of an urban settlement, neighbourhood types (residential, industrial etc.), or building types (detached, semi-detached, terrace etc.).

One area that could obviously benefit of richer semantics in spatial databases is map generalisation. Map generalisation aims to derive a model of the geographic reality that is appropriate for portrayal at a certain scale and purpose. It is important to note that this abstraction process is not just a matter of simplification of detailed situations to reduce spatial clutter and therefore guarantee legibility of a map; rather, different phenomena and patterns have to be portrayed at various scale levels (Brassel & Weibel, 1988). Bertin (1967/1999) therefore distinguishes *conceptual generalisation* and *structural generalisation*. Conceptual generalisation happens when "a city emerges from a collection of houses and streets", or a "coal pan from a collection of coal mines". Structural generalisation simplifies geometry, but conserves conceptualisation. More recently, this dichotomy has been termed *model (or model-oriented) generalisation* and *cartographic generalisation* (Grünreich, 1992).

While higher level geographic concepts are not explicitly coded in current spatial databases, they are nevertheless *implicitly* contained, owing to the fact that there often exists a relationship between the form (i.e. geometry) and function (i.e. semantics) of real-world phenomena, particularly in the built environment. Hence, it is possible – at least to some extent – to 'enrich' spatial databases retrospectively, making implicitly contained higher level geographic concepts explicit. This process is termed spatial database enrichment.

In particular, spatial patterns in the urban domain provide the basis for a variety of applications, such as urban planning or

* Corresponding author. Tel.: +41 44 635 52 17; fax: +41 44 635 68 48.
E-mail addresses: patrick.luescher@geo.uzh.ch (P. Lüscher), robert.weibel@geo.uzh.ch (R. Weibel), dirk.burghardt@tu-dresden.de (D. Burghardt).

0198-9715/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2009.07.005

pedestrian navigation (Lüscher, Weibel, & Mackaness, 2008). The obvious example, again, is map generalisation. Take the case of a building that is too small to be fully legible on a target map. Here, semantic information is useful in deciding how to proceed: If the building is in a rural area (and hence rather isolated and presumably important), the building may be slightly enlarged; if it is in an urban area, it may be eliminated; and if it happens to be a special type of building such as a hospital, it may be replaced by a special symbol (Steiniger, 2007).

While there are a number of specific algorithms for data enrichment in spatial databases (Lüscher et al., 2008), the goal of the work in the present paper is to provide a modular approach to the overall process. The definition of spatial patterns is formalised through ontologies, which in turn can be used to drive the pattern recognition process.

The general approach was presented in an earlier paper (Lüscher et al., 2008). In the present paper, the following research questions are covered:

1. What methods are suited to classify instances with respect to formal definitions?
2. To what extent is it possible to use only simple measures (such as area and topological relations) to define complex concepts?

The premise is that the pattern recognition process needs to respect uncertainty of spatial data and vagueness of spatial knowledge. To address the first research question, an approach is presented that translates the ontology into a Bayesian network for carrying out fuzzy inference and for including training data. The approach is illustrated step-by-step using a case study that classifies English terraced houses in a topographic dataset. To address the second research question and to put the approach into the context of previous attempts to formalise pattern recognition, an alternative ontology that avoids complex spatial measures is taken as reference. Both ontologies are used to classify four English urban areas.

The remainder of this paper is organised as follows. Section 2 reviews previous approaches to model-based spatial pattern recognition. In Section 3 the need for ontology-driven pattern recognition process is presented, and the approach is outlined. In Section 4 we introduce the case study used in this paper – terraced houses – and define the corresponding ontology. Section 5 argues for an approach of fuzzy inference, based on the translation of the ontology into a Bayesian network. Section 6 presents two sets of experiments, one using a basic ontology not specifically defined for spatial database enrichment, and a second one using the ontology as developed in Section 4. Section 7 presents classification results. Section 8 discusses the ontology-driven approach with particular emphasis on the comparison of the two experiments. Finally, Section 9 rounds off the paper by conclusions and an outlook on future research.

2. Review of relevant literature

2.1. Related work on ontology-based spatial pattern recognition

Klien (2007) presents a framework for annotation of geodata, using Semantic Web technologies (Yu, 2007). She defines semantic annotation as creating links between feature types of a dataset and concepts of an external ontology, and argues that linking based on string-similarity of type/class names alone is too inaccurate. The semantic descriptions in the ontology are therefore used to derive instances of concepts and compare them with actual instances in the database. For example, she defines `flood plain` as a flat area adjacent to a river and not very much higher in altitude than the river (such that the area is regularly subject to flooding). This definition is translated to the Semantic Web Rule Language (SWRL, 2009). Spatial relations (such as `adjacent`) are mapped to spatial analysis operations, and regions representing flood plains are inferred through logic deduction. She argues that by following this strategy, instead of implementing a 'black box'-approach, increased flexibility and transparency to the user is achieved. However, it is further argued that automatic classifications produced by the method are likely error-prone and need to be presented to a human user for final confirmation.

Thomson and Béra (2008) present a methodology for generating urban residential land-use through logic deduction. Increasingly complex spatial aggregates are generated starting from atomic concepts like house, garden, or road. As in the work of Klien (2007), spatial predicates are generated through spatial analysis operations in a GIS and exported to OWL–DL. The Web Ontology Language (OWL, 2008) is a family of languages to author ontologies. Classification of buildings and plots is then carried out through Description Logic subsumption reasoning (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2003).

Zhang, Stoter, and Ai (2008) propose a similar approach, although their goal is to improve reusability in cartographic constraint evaluation. During cartographic generalisation, cartographic constraints describe particular spatial settings for which preferred actions exist. For example, there exists a constraint that specifies that 'roads leading to an isolated building should not be omitted'. Hence, detecting spatial settings corresponds to spatial pattern recognition. The proposed approach works again by decomposing complex spatial settings into simpler measures, and use some kind of predicate logic and/or terminological reasoning to infer instances, although a more detailed account of implementation is not given.

2.2. Uncertainty of geographic objects

Many concepts in the geospatial domain are poorly defined and traditional crisp logic is insufficient in dealing with uncertainty. Klien (2007) points out that "the notion 'relatively low' is not expressible in the logic of the representation language" (p. 444), but does not consider uncertainty in her framework further. According to Fisher (1999), there are two kinds of uncertainty associated with poorly defined concepts:

- *Vagueness*, which arises from poor definition of a class or individual object. As a consequence of vagueness, the extent of many spatial phenomena cannot be delimited sharply.
- *Ambiguity*, which arises from differing classification systems. The same road could be denoted as Expressway (by someone with US American background) or as Motorway (by someone with British background).

Dissolving ambiguity for enabling interoperability is one of the main applications of ontologies (Agarwal, 2005). Often concepts do not map one-to-one, but their meaning overlaps partially. Hence, there is increasing research interest in extending conventional reasoning with probabilistic techniques such that not only identical concepts can be deduced, but the most similar ones (Sen, 2008). Translating traditional OWL representations to Bayesian networks (Russel & Norvig, 2003) to carry out probabilistic reasoning is a promising approach (Zheng, Kang, & Kim, 2007). Recently, extensions such as PR-OWL (Costa & Laskey, 2006) or BayesOWL (Ding, Peng, & Pan, 2006) have been introduced to formalise translations from OWL into Bayesian networks.

Vagueness in classification arises because realisations of concepts are often imperfect and come with certain variations. For instance, ponds can be defined as a water body smaller than a lake, but the transition from pond to lake is gradual. As a consequence we are unable to define crisp thresholds for class membership. Fuzzy set theory (Fisher, Wood, & Cheng, 2004; Ladner, Petry, & Cobb, 2003) is an approach to account for this kind of uncertainty by defining fuzzy memberships. An alternative approach is Bayesian decision theory, by which class membership probabilities are estimated.

2.3. Contributions

The key contribution of this paper is the combination of an approach for ontology-based spatial pattern recognition with probabilistic inference to account for vagueness. A probabilistic Bayesian approach is used for inference. The advantages of Bayesian inference are discussed in Section 5 and can be summarised as follows:

- Good integration into ontologies as shown by previous work on probabilistic OWL;
- sound inference also when multiple decisions are chained; and
- the simplicity of learning conditional probabilities from training data.

A second contribution is the introduction of *abstract concepts* that are defined algorithmically, but are formulated as simply and generally as possible (so that they can be re-used). A third contribution is the evaluation of the robustness of ontology-driven spatial database enrichment using large extracts of real data.

3. Ontology-driven spatial database enrichment

Lüscher et al. (2008) discussed algorithmic approaches to spatial database enrichment and argued why ontologies should be used to drive the pattern recognition process. Undoubtedly, existing algorithmic methods have been successful in detecting specific spatial patterns, but solutions that solely rely on algorithms also exhibit several important weaknesses:

- They have often been developed and parameterised for specific data models and databases. That *limits the reusability* of pattern recognition methods across different databases.
- They often make use of bespoke geometric algorithms and/or statistical techniques that do not reveal the 'mechanics' of the recognition procedure. Hence, they have *limited transparency* and explanatory value for the end user.
- They typically cannot be adapted to take into account additional information in the detection procedure, such as topography, which may be important in describing the genesis of certain patterns. That is, they have *limited extensibility*.

Ontologies have the potential to better inform the pattern recognition process with the aim of improving on some of the limitations of purely algorithmic approaches. Spatial concepts and their (spatial) relationships to other, 'lower level' concepts are explicitly modelled in an ontology. While the lowest level concepts are extracted through traditional spatial pattern recognition processes, they can be used to infer the existence of higher level concepts.

This ontology-driven approach proceeds in four steps (Lüscher et al., 2008): We draw on textual descriptions of urban spaces (step 1), then formalise these patterns, their context and hierarchical composition using methods from ontological engineering (Gómez-Pérez, Fernández-López, & Corcho, 2003) (step 2). The ontological definitions of patterns are then used to deductively trigger appropriate pattern recognition algorithms (step 3) in order to detect them in real spatial databases (step 4).

We use the term 'ontology' in the sense of the engineering sciences, where it is usually defined as an explicit specification of a shared conceptualisation (Gruber, 1993). It is thus an attempt to capture the knowledge of a certain domain in a systematic way by breaking it down into the types of entities (*concepts*) that exist and the *relations* that hold between them. Therefore, in a first step, knowledge about the domain has to be collected. In this study, knowledge was extracted from the literature on urban development and urban history, complementing this information with the help of dictionaries and thesauri.

4. Ontologies of urban space descriptions

4.1. The case study of English terraced houses

It should be noted that according to the ontology definition given by Gruber (1993), there can be multiple ontologies for the same concept depending on the purpose the ontology is modelled for. The purpose of this research is to model ontologies for the detection of geographical concepts in spatial databases. Such an ontology has been built for the extraction of terraced houses (also called terrace houses or terraces) as they are conceptualised in *urban morphology*. Relevant concepts of the domain were extracted from a thesaurus of urban morphology (Jones & Larkham, 1991). Several case studies (e.g. Conzen, 1969) and a compendium about "The English Terraced House" (Muthesius, 1982) then gave more insight in the understanding of the concepts. By way of example, Fig. 1 shows residential house types identified in the urban morphology literature. Mappings of terraced house settlements are provided in Section 7.

We use terraced houses as a case study for several reasons. First, they represent the most widespread housing type in English cities (Muthesius, 1982) and building types such as terraced, semi-detached, and detached houses are commonly used in everyday speech. For instance, they give essential clues to prospective house buyers as to what to expect when reading through real estate advertisements (King, 1994). Second, knowledge about terraces, semi-detached and detached houses is also important in map generalisation. House types are used for typification of residential plots; for example, yards are merged differently in terraced house settlements than in detached and semi-detached settlements. Third, the concept of the terraced house integrates various low level concepts (as will be shown below) that can be re-used in similar concepts (e.g. other residential house types). And finally, it forms in turn a low level concept of other high level concepts, such as 'residential area'. Hence, it may serve as an exemplar for testing the versatility and reusability of the ontology-driven approach to spatial database enrichment.

A textual description of the English terraced house can be summarised as follows: The construction of terraced houses is closely linked to the Public Health Act of 1875, which was established to improve urban living conditions and resulted in re-housing of population from slum clearance areas (Conzen, 1969). The demand for cheap mass housing was met by creating rows of unified buildings sharing sidewalls. Owing to the low social status of the original

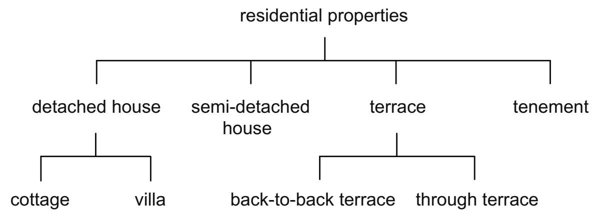


Fig. 1. Urban residential house types extracted from the "Glossary of Urban Form" (Jones & Larkham, 1991).

dwellers, lot sizes and room footprints are small. Terraced houses usually have small front-gardens and possibly attached sculleries and a yard at the rear. Often, multiple rows of houses form an area of a highly regular plot pattern.

A concept map constructed from these descriptions is shown in Fig. 2. Relations to simple properties, such as the area of a polygon, were included into the box of the concept itself, while relations that connect two (or more) concepts are drawn as arrows between them. This is for clearer visualisation only.

In the figure, `terraced house` is defined by its relations to other concepts. Some of those concepts are defined by relating them to even more basic concepts. For instance, the Oxford English Dictionary (Simpson & Weiner, 1989) defines a yard as "a comparatively small uncultivated area attached to a house or other building, or enclosed by it". This means, `yard` is defined by its area and its relations to `uncultivated area` and `building`. The concept map also contains *abstract concepts* which are to be implemented algorithmically as they constitute general units that are inefficient to break up further. One example is the concept `row of houses`, which denotes a linear, homogeneous arrangement of adjacent houses.

Having modelled terraced houses *as conceptualised by humans*, the concept map must be formalised to a pattern recognition process. This consists of two steps: On the one hand, explicit semantics have to be assigned to abstract concepts and relations by mapping them to (often spatial) operations. On the other hand, an algorithm has to carry out the classification process, inferring instances of concepts defined in the ontology. Through these steps, an ontology is defined. In the remainder of this section, mapping of relations and concepts is discussed. Section 5 presents an approach for fuzzy inference, based on the translation of the ontology into a Bayesian network.

4.2. Mapping of spatial relations and abstract concepts

The meaning of predicates such as `adjacentTo`, `presenceOf`, and `hasArea` has to be interpreted by spatial analysis. `adjacentTo` denotes topological connection (i.e. adjacency) of two areas. The custom of embedding residential houses between front yards and backyards leads to a high proportion of green space in residential settlements. This can be used to establish a contextual measure whether a house lies in a residential neighbourhood or not. `presenceOf(yards)` was therefore mapped to a kernel density measure as it was developed by Chaudhry and Mackaness (2008). Yard density at any location k is given by:

$$yd_k = \sum_{i=1}^{n} \frac{\sqrt{a_i}}{d_{ki}^2} \qquad (1)$$

where a is the area of yard i, d_{ki} the distance between location k and yard i, and n is the number of yards involved in the calculation of density.

It was also mentioned above that some concepts were left abstract because it is inefficient or impossible to define them by relations alone. These involve custom-built algorithms for their instantiation. For the terraced houses ontology, this had to be done for `row of houses` and `areas of parallel rows`. The algorithms are discussed in full detail in Lüscher et al. (2008) and are only briefly sketched here. Perceptual alignments were obtained by grouping buildings sharing a common wall and then connecting the centroids of the buildings to a path. The path was broken up at sharp turns, i.e. where the angle between two consecutive segments was larger than 60°. Remaining groups were finally qualified for homogeneity and straightness. The concept `areas of parallel rows` was derived by identifying the main axes of building groups, clustering these groups using the direction of the axes, and finally qualifying clusters for their homogeneity.

5. Bayesian inference as a technique to derive instances of concepts

5.1. Bayesian inference as a means to integrate probabilistic and crisp decisions

Bayesian inference is a standard approach in pattern classification (Duda, Hart, & Stork, 2001; Rice, 1988; Russel & Norvig, 2003). Assume that we have a categorical variable C that is statistically dependent on a set of evidence variables $F_1, \ldots, F_n$. For instance, C could be a binary variable that describes the fact whether a building constitutes a terraced house or not, depending on whether it is contained in a homogeneous alignment of houses, the presence of yards, etc.

The Bayesian decision rule tries to minimize the probability of error in a decision by deciding for the most probable outcome. Consider Fig. 3, which shows a hypothetical likelihood curve for

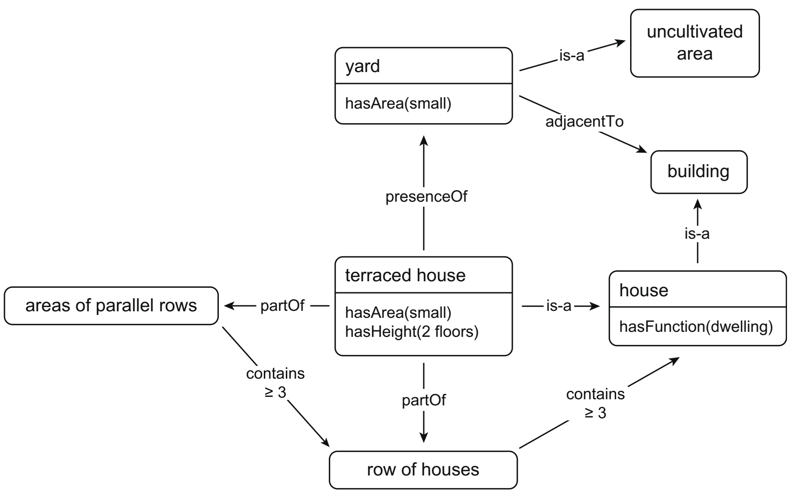


Fig. 2. A concept map of terraced houses suited for data enrichment.

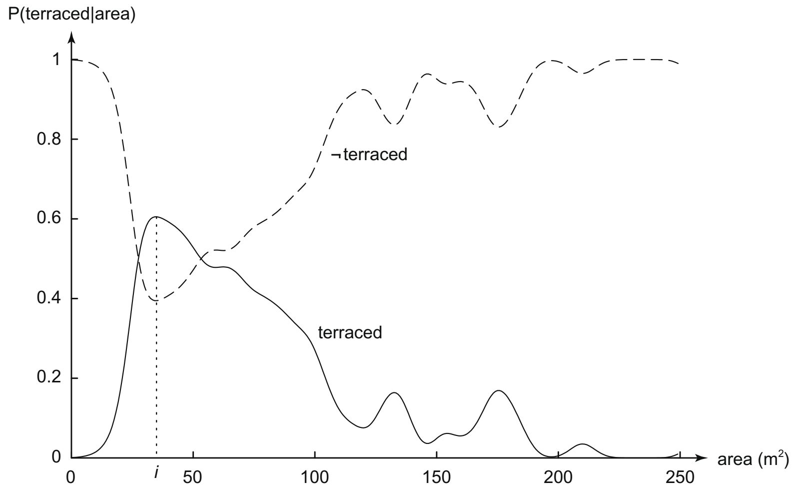


Fig. 3. Hypothetical likelihood curve for terraced house, given the building area.

a building to be a terraced house, if the decision was based exclusively on its area. Let's assume building i having area 35 m^2 has to be classified. The likelihood of being 'terraced' as indicated in the figure is 0.6, while the likelihood of being 'not terraced' is only 0.4. Therefore we decide building i is 'terraced'.

Formally, the Bayesian decision rule states that the predicted class $\hat{C}$ for a given realisation $F_1 = f_1, \ldots, F_n = f_n$ is the class c which maximises the *likelihood* $P(c|f)$. This is mathematically expressed using the operator arg max:

$$\hat{C} = \arg\max_c P(C = c | F_1 = f_1 \wedge \ldots \wedge F_n = f_n) \tag{2}$$

Any inference can be translated into a conditional probability, including crisp relations with Boolean outcomes, as it happens when an *is–a* relation is turned into a Bayesian decision. The likelihood distribution is trivial in these cases, as shown in Table 1.

In the general case, if there are more evidence variables involved than just one, the evidence variables are usually not independent of each other. That is, a joint likelihood distribution has to be created upon which the Bayesian decision is based.

5.2. Chaining Bayesian decisions

The inference process starts with the concepts that can be derived using only concepts that are already in the database, and proceeds incrementally to derived concepts of higher order. In this manner the inference task is translated into a chain of Bayesian decisions, creating a so-called Bayesian network. Probabilistic inference in Bayesian networks is theoretically well explored (Russel & Norvig, 2003). Consequently, the ontology is turned into a Bayesian network by specifying joint conditional probability distributions for each concept. This can be trivial as in the case of the *is–a* relation. When fuzzy relations are involved such as in the example of building area, it is easier to learn probability distributions from training samples instead of specifying them manually. In the following section we will show how this can be achieved.

5.3. Learning Bayesian decisions from training data

If the likelihood is to be learned from training data, Eq. (2) can be transformed to a more convenient form. The transformation makes use of the Bayes' theorem:

$$P(C = c | F_1 = f_1 \wedge \ldots \wedge F_n = f_n) = \frac{P(F_1 = f_1 \wedge \ldots \wedge F_n = f_n | C = c) P(C = c)}{P(F_1 = f_1 \wedge \ldots \wedge F_n = f_n)} \tag{3}$$

The denominator on the right hand side is a scaling factor that guarantees that probabilities sum to one. Recalling the Bayesian decision rule and Eq. (2), we are only interested in for which value of c the term on the right hand side reaches its maximum. The denominator is independent of c and can therefore be omitted, leading to the following formulation of the Bayesian decision:

$$\hat{C} = \arg\max_c P(F_1 = f_1 \wedge \ldots \wedge F_n = f_n | C = c) P(C = c) \tag{4}$$

In Eq. (4), likelihood has been replaced by the *class-conditional joint probability density function*.

The advantage of Eq. (4) is that density distributions can be estimated using training data. A convenient method to estimate them is to employ kernel density estimation (Silverman, 1986). One can guarantee that the probabilities sum to one if a standard normal distribution function is chosen as kernel. Let $\vec{f} = (f_1, \ldots, f_n)$, where $\vec{f}_i$ are the training samples with classification $C = c$. The joint conditional density distribution P_c is then given by:

$$P_c(\vec{f} | C = c) = \frac{1}{N \|\vec{h}\|} \sum_{i=1}^{N} K\left(\frac{\vec{f} - \vec{f}_i}{\vec{h}}\right), \quad \text{where } K(\vec{x}) = \frac{1}{(2\pi)^{N/2}} e^{-0.5 * \vec{x}^T \vec{x}} \tag{5}$$

N is the number of samples and $\vec{h}$ are the bandwidths, which constitute smoothing factors for the density function.

Fig. 4 illustrates the calculation of the probability density function. The crosses below the x-axis indicate building area values of terraced houses that were tagged in the training data. Dashed lines are kernels for each sample. The solid line indicates the estimated density function, which is the sum of individual Gaussian curves.

6. Experiments

Two experiments were carried out: The first experiment is based on a definition of `terraced house` by the Ordnance Survey (Ordnance Survey Ontologies, 2008). This experiment was designed to provide a reference of how much prediction accuracy could be achieved by compact definitions and crisp logic reasoning

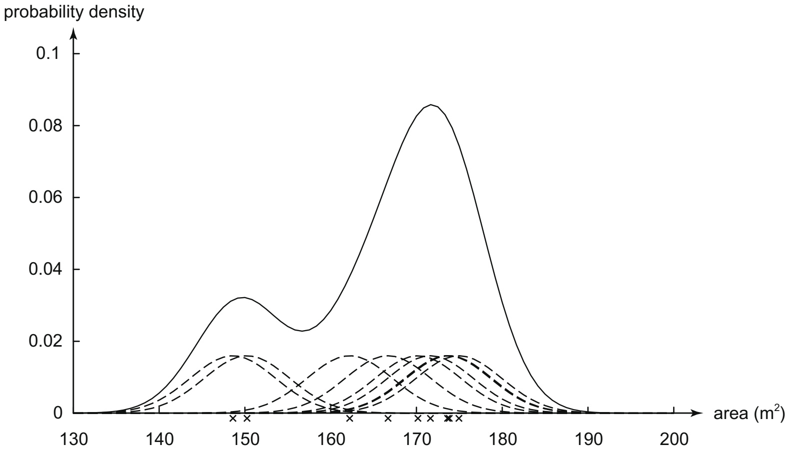


Fig. 4. Illustration of density estimation using a Gaussian kernel.

alone, and where typical problems would arise. In the second experiment, test areas were classified according to the Bayesian inference approach presented above. In Sections 7 and 8, we evaluate classifications by means of their prediction accuracy (compared to human interpretation) and show some typical errors for both experiments.

6.1. Test data

Four urban areas of England were extracted from the Ordnance Survey MasterMap® Topography Layer for the cities of Middlesbrough, Norwich, Portsmouth, and Southampton. The OS MasterMap® Topography Layer models topographic features in urban areas corresponding to a scale of approximately 1:1250. The extents of the test datasets were chosen such that they include not only residential areas, but a wide variety of urban land-use, i.e. mixed residential with smaller commercial buildings, and large industrial/commercial grounds. Besides traditional, Victorian and Georgian-type residential areas, also more recent settlements were found in the areas. They differ from the traditional type in a less regular arrangement, and a mix of terraced and semi-detached housing types in one block (Marshall, 2005). In the experiments no distinction was made between the two types of settlement periods.

The authors manually attributed buildings in all datasets with 'terraced'/'not terraced' by visual inspection. Besides MasterMap®, aerial photographs provided by Google Earth were used for the manual classification. Table 2 shows some characteristics of the study areas.

The data enrichment process starts from concepts that are readily available in the database. For instance, for `building`, an attribute in OS MasterMap® encodes whether a polygon represents open land, transportation or a building. Likewise, instances of `uncultivated area` can be defined through a combination of two attributes. Therefore, relations were added to the ontologies that define buildings and uncultivated areas as presented in Table 3 (in SWRL Human Readable Syntax).

6.2. Experiment based on simple ontology

This experiment was carried out to reveal insights to which extent classification is possible based on very basic spatial operations and crisp inference.

Table 1
Probability distribution for an *is–a* relation.

P (house)	is–a (building)
1.0	True
0.0	False

Table 2
Characteristics of the study areas.

Study area	Area covered (east–west/ north–south) (km)	# Buildings	# Terraced houses in manual classification
Middlesbrough	3.33/3.26	41,667	14,138 (33.9%)
Norwich	5.63/4.75	62,021	20,297 (32.7%)
Portsmouth	5.80/6.55	80,853	37,862 (46.8%)
Southampton	3.85/2.80	22,950	5075 (22.1%)

Table 4
Rabbit definition of terraced house as provided by the Ordnance Survey.

`House`	Every `House` is a kind of `Building`. Every `House` has purpose Housing of People.
`End Terrace House`	An `End Terrace House` is anything that: • is a kind of `House`; • is connected to exactly one `Terrace House`.
`Terrace House`	A `Terrace House` is anything that: • is a kind of `House`; • is connected to exactly 2 `Terrace Houses`; or is connected to exactly one `End Terrace House` and is connected to exactly one `Terrace House`.

Table 3
Rules for asserting `building` and `uncultivated area` from OS MasterMap®.

ArealPrimitive(?x) ∧ hasAttribute(?x, ?y) ∧ hasName(?y, "DescGroup") ∧ hasValue(?y, "Building") ⇒ Building(?x)
ArealPrimitive(?x) ∧ hasAttribute(?x, ?y) ∧ hasName(?y, "DescGroup") ∧ hasValue(?y, "Building") ∧ hasAttribute(?x, ?z) ∧ hasName(?z, "make") ∧ hasValue(?y, "multiple") ⇒ UncultivatedArea(?x)

Table 5
Rules for classifying terraced houses used in the first experiment.

Building(?x) ∧ hasArea(?x, ?a) ∧ swrlb:greaterThanOrEqual(?a, 35) ∧ swrlb:smallerThanOrEqual(?a, 160) ⇒ House(?x)
Building(?x) ∧ isConnectedTo(?x, ?y) ∧ House(?y) ⇒ isConnectedToHouse(?x, ?y)
House(?x) ∧ (=2 isConnectedToHouse)(?x) ∧ ⇒ MidTerracedHouse(?x)
Building(?x) ∧ isConnectedTo(?x, ?y) ∧ MidTerracedHouse(?y) ⇒ isConnectedToMidTerracedHouse(?x, ?y)
House(?x) ∧ (=1 isConnectedToMidTerracedHouse)(?x) ⇒ EndTerracedHouse(?x)

Table 6
Exemplary OWL code for a building.

```
<Building rdf:about="http://www.geo.uzh.ch/orus#osgb1000002054799448">
<hasArea>55.35714999958873</hasArea>
<isConnectedTo rdf:resource="http://www.geo.uzh.ch/
    orus#osgb1000002054799449"/>
<isConnectedTo rdf:resource="http://www.geo.uzh.ch/
    orus#osgb1000002054949238"/>
<isConnectedTo rdf:resource="http://www.geo.uzh.ch/
    orus#osgb1000002054949232"/>
<isConnectedTo rdf:resource="http://www.geo.uzh.ch/
    orus#osgb1000002054799458"/>
</Building>
```

The Ordnance Survey GeoSemantics team provides ontologies of their spatial databases (Ordnance Survey Ontologies, 2008). The aim is to describe the content of OS databases concisely to improve usability and data integration. The first classification experiment was based on the description of terraced house provided in the 'OS ontology for Buildings and Places'. The natural language description is as follows: "A terrace house is one that is part of a line of connected houses" (Ordnance Survey Ontologies, 2008). The Ordnance Survey GeoSemantics team provides equivalent definitions in Rabbit, a controlled language for authoring ontologies (Hart, Johnson, & Dolbear, 2008), and OWL. Table 4 shows the Rabbit definition to ease reading.

The definition differentiates between houses at the end of a terrace (`End Terrace House`) and houses within a terrace (`Terrace House`). We will denote the latter type `Mid Terrace House` to make a clear distinction. The definition is based on only two types of relations: the functional definition `hasPurpose(Housing)`, and the topological relation `isConnectedTo()`.

In order to carry out the reasoning, the original Ordnance Survey definition was modified in two points. Firstly, there was no information available in OS MasterMap® whether a building serves for dwelling. One possibility would be to integrate data that provide missing information (e.g. from zoning maps or a building register). However, this option was not pursued in this study, as the focus was on pattern recognition from a single topographic database. The `hasPurpose(Housing)` relation was therefore replaced by a restriction on the area of the building footprint as an approximation. The cut-off values were determined experimentally.

Secondly, the rule for `Mid Terrace House` contains a reference to itself, which makes reasoning unfeasible by the forward-chaining reasoning mechanism that was employed in the experiment. The rule was therefore simplified to "is connected to exactly 2 Houses". The modified rules used for classification are given in Table 5.

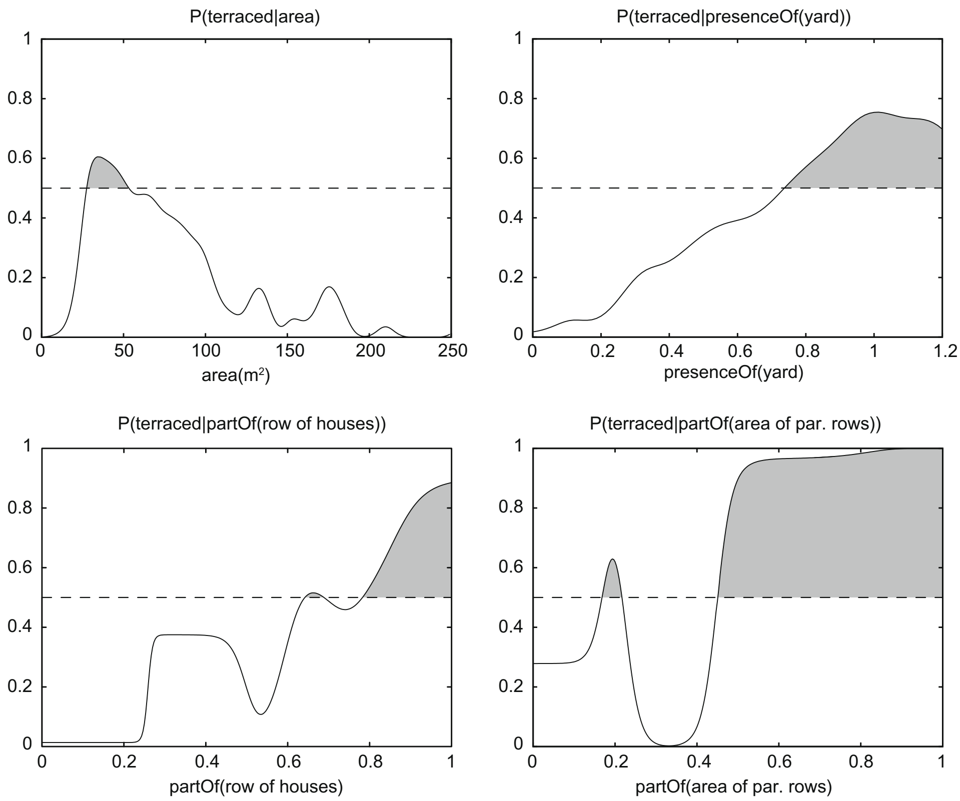


Fig. 5. Marginal probability distributions for uncertain relations of the terraced house concept. Shaded grey areas: regions with $P(\text{terraced}) > 0.5$.

Table 7
Comparison of classification produced by the simple ontology approach and human interpretation.

Area	# Buildings	# Correct class	% Correct class	Precision (%)	Recall (%)	Cohen's kappa
Southampton	22,950	17,223	75.0	77.7	86.4	0.76
Middlesbrough	41,667	37,919	91.0	94.8	77.8	0.79
Norwich	62,021	55,306	89.2	88.8	76.2	0.74
Portsmouth	80,853	75 ,559	93.5	92.5	93.6	0.87

Table 8
Comparison of classifications produced by Bayesian inference and human interpretation.

Area	# Buildings	# Correct class	% Correct class	Precision (%)	Recall (%)	Cohen's kappa
Southampton[a]	22,950	22,283	97.1	90.7	96.8	0.92
Middlesbrough	41,667	40,899	98.2	98.3	96.3	0.96
Norwich	62,021	59,388	95.8	93.1	94.0	0.90
Portsmouth	80,853	76,942	95.2	92.8	97.3	0.90

[a] Terraced houses of the Southampton area were part of the training sample.

The Jena general purpose reasoning engine (Jena, 2009) was employed to carry out reasoning. Data exchange between the spatial database and Jena happened through OWL as exchange format: For each building, a Java program calculated area and topological connectedness to other buildings, and added this information as OWL properties. As an example, Table 6 shows an OWL extract for one building.

The reasoner thereon classified terraced houses according to the rules presented in Table 5. The classifications were transferred back into the GIS for controlling the results.

6.3. Experiment based on Bayesian approach

This experiment used the ontology as presented in Fig. 2. The Ordnance Survey MasterMap® datasets used in this study did not have an attribute for the number of floors of buildings. As mentioned previously, there was also no information about building function available. Therefore `house` and `hasHeight(2 floors)` were dropped and the relations pointing to `house` were short cut to `building`. The authors assessed that the remaining criteria `hasArea(small)` and `presenceOf(yards)` provide in most cases enough discriminatory power for classification.

As in the previous experiment the ontology was stated as a set of rules, but inference was carried out using a Bayesian reasoner, which was implemented as a custom-built prototype for ontology-driven database enrichment in Java. Whenever the reasoner has to check if a database object is an instance of a concept, it calls analysis routines for each predicate in the definition of the concept. The routines implement necessary (spatial) analysis functions. Fuzzy predicates are allowed to return any number. For instance, `hasArea(small)` returns the footprint of the database object (e.g. building); `presenceOf(yards)` returns the density of yards at the location of the object. The obtained values constitute the evidence variables for Bayesian inference. The Bayesian inference uses

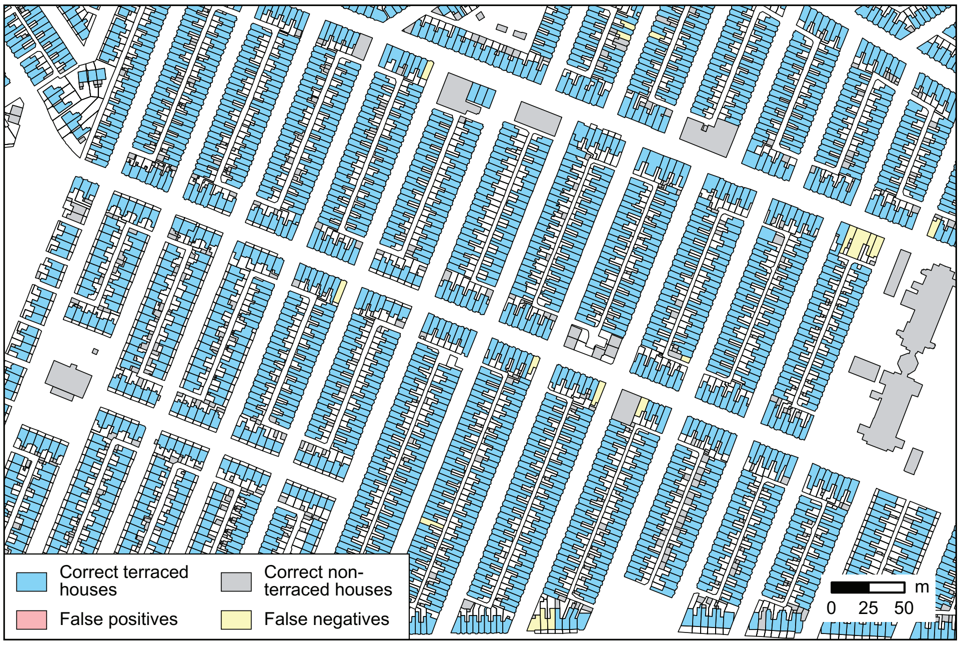


Fig. 6. Traditional terraced house neighbourhood (Middlesbrough). Please note that there are no false positives in this area. OS MasterMap data Ordnance Survey ©Crown Copyright. All rights reserved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Modern terraced neighbourhood (Norwich). Colouring as in Fig. 6. OS MasterMap data Ordnance Survey ©Crown Copyright. All rights reserved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

training data for estimating a joint probability density distribution as explained in Section 5.3. For each concept definition having fuzzy predicates, a set of positive and negative examples must therefore be given.

For the `terraced house` concept, all 5075 buildings of the Southampton area tagged as *terraced house* were selected as positive samples. A characteristic set of 6629 buildings from the Southampton area was selected to form samples of non-terraced houses. Fig. 5 shows the marginal probability distributions derived from these sample data. Grey shaded areas denote the acceptance of terraced house, if the decision was based on one criterion only. The *is–a (Building)* predicate in the definition of `terraced house` is crisp and has a probability distribution as shown in Table 1.

7. Evaluation of classification accuracy

In the following, the classification accuracy of the conducted experiments is measured statistically by comparison to human interpretation. Classification accuracy was measured by means of *precision*, *recall,* and *Cohen's kappa coefficient* κ. *Precision* indicates the probability that a terraced house found by a classification algorithm was also classified as terraced house in manual classification. *Recall* indicates the probability that a manually classified terraced house is found by the classification algorithm. *Cohen's kappa* (Lillesand, Kiefer, & Chipman, 2000) is a measure of agreement between classifications; $-1 \leqslant \kappa \leqslant 1$, whereby high values of κ denote a good agreement.

Table 7 presents results produced by the simple ontology approach. The Portsmouth area is classified very well, while results of the other three study areas produce a lower kappa value. This is explained by the fact that the Portsmouth area is a 'standard' situation in the sense that highly regular terraced houses dominate. Pure residential areas were classified generally well, while accuracy in mixed-use and industrial areas was lower.

Table 8 presents the classification accuracies for the experiment using Bayesian inference. It shows that high classification accuracy could be achieved in all four study areas.

Fig. 6 shows a traditional terraced house neighbourhood as classified by the Bayesian inference approach. Fig. 7 depicts a situation in a more 'modern' type of settlement, having lower building density and less stringent regularity of the arrangement of rows.

8. Discussion

In the following, the Bayesian approach is assessed by comparison to the more traditional simple ontology approach and by making considerations on scalability. The benefits are clarified by means of relating the approach to the case study. Finally, we conclude with perspectives for future research.

8.1. Comparison of common errors produced by the approaches

In the following, we contrast both approaches by discussing common sources of disagreement between the human interpretation and automatic classification as produced by each approach.

8.1.1. Common errors produced by the simple ontology approach

Errors produced by the simple ontology approach can be grouped into two classes.

Missing linear arrangement: Fig. 8a shows a case where porch roofs classified as house prevent correct classification of a terraced house. The house indicated as 'MT' was classified as mid-terraced because it connects to exactly two other houses (one of them being the incorrectly classified porch roof). Buildings indicated as 'ET' were classified as end-terraced, because they connect to the mid-terraced house. Most of the terraced houses were not found; they connect to more than two other houses (including the porch roofs). Fig. 8b depicts a situation where terraced houses were produced in a heterogeneous, dense built-up block. Even if the buildings in the situation constitute dwellings, the situation would not be perceived as row of terraced houses, but as an assembly of houses randomly built together. The errors in both situations occur because topology alone does not capture the fact of being 'a line of houses'. A synoptic view is needed to decide on what constitutes an align-

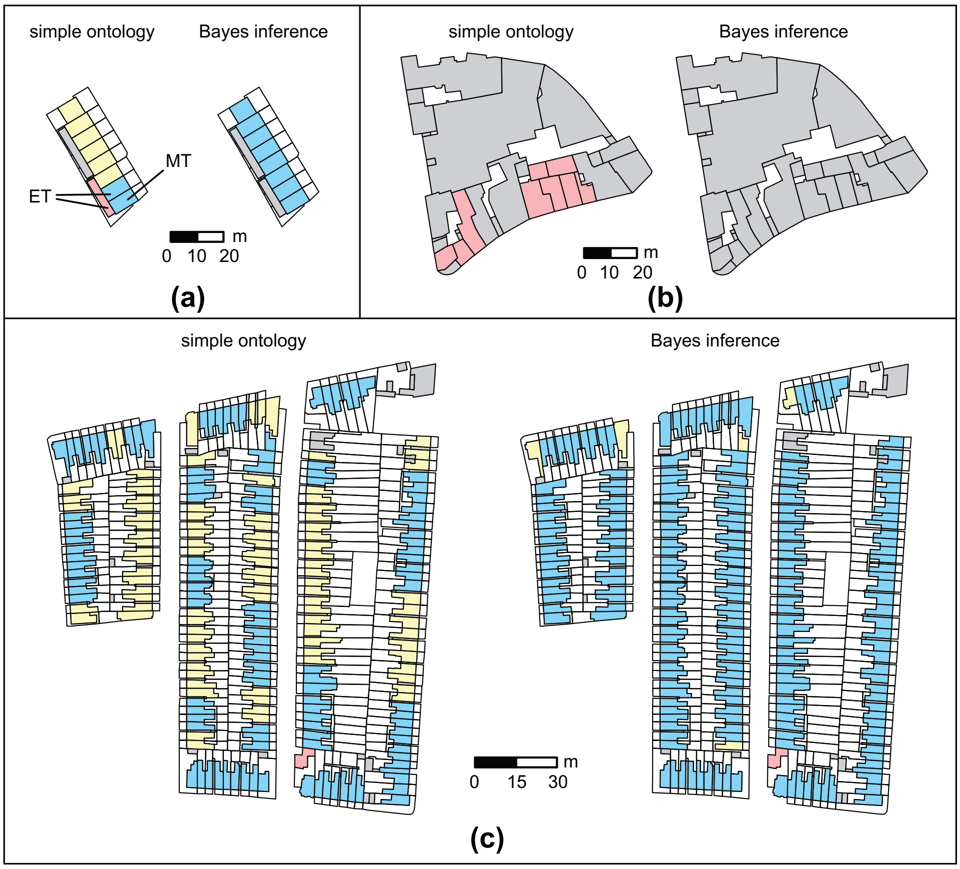


Fig. 8. Typical errors produced by the simple ontology approach. Colouring as in Fig. 6. OS MasterMap data Ordnance Survey ©Crown Copyright. All rights reserved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ment and which houses are parts thereof. In the Bayesian inference experiment this synoptic view is provided by the `row of houses` concept.

Special cases in the modelling of features: Fig. 8c shows a situation where small, oblong polygons disconnect otherwise perfectly regular terraces. The polygons are actually small enclosed alleys that connect the street to the backyards. Although they are integrated into the houses (e.g., the first floor above the alley is made up of a room), they are modelled as separate polygons in OS MasterMap®. As a consequence, houses are not topologically connected and are not detected as terraced houses. In the Bayesian inference experiment, the `row of houses` concept again provides grounds for correct classification. Here, further experiments are needed to establish whether such special cases can be modelled in SWRL rules. However, this would in turn render the description less compact.

Concluding, the simple ontology approach produced reasonable results where situations corresponded to the prototypical conceptualisation. In less clear situations, a synoptic view is missing that cannot be constructed using logic reasoning alone.

8.1.2. Common errors produced by the Bayesian inference approach

A type of *false positive* produced by the Bayesian inference approach is shown in Fig. 9a. There are rows of garages or sheds in the backyards having an area of around 30 m^2. These were classified as terraced houses by the Bayesian inference approach. The simple ontology experiment, applying a higher area threshold of 35 m^2, did not reproduce this behaviour, but missed terraced houses in the leftmost vertical row that have an area below 35 m^2. Obviously, this issue of features with overlapping values cannot be solved without adding more criteria (e.g., detecting backyard sheds in advance).

A second but infrequent type of false positive is shown in Fig. 9b. It shows rows of semi-detached houses that are connected with each other through small constructions such as shelter roofs at the entrance. The automatic classification treats them as terraced houses, although they rather correspond to semi-detached houses because effectively they have three exterior walls that can provide more daylight to inhabitants, whereas terraced houses only have two walls (excluding houses at the end of terraces). In this specific case, the simple ontology approach did not show this misclassification due to the modelling discussed already discussed in Fig. 8c (hence also the many false negatives).

False negatives were less frequent than false positives. Typically they occurred at boundaries of residential areas, along large streets, and in isolated terraces, where `presenceOf(yard)` was generally lower.

A general source of disagreement arose in some cases when rows of buildings were not discernable from terraced houses in MasterMap® alone, but from information that was only visible in aerial photographs, such as facades and patterns of access paths and entrances. In other cases, the human operator judged buildings to have a different function than dwelling based on the spatial context visible in the aerial photograph.

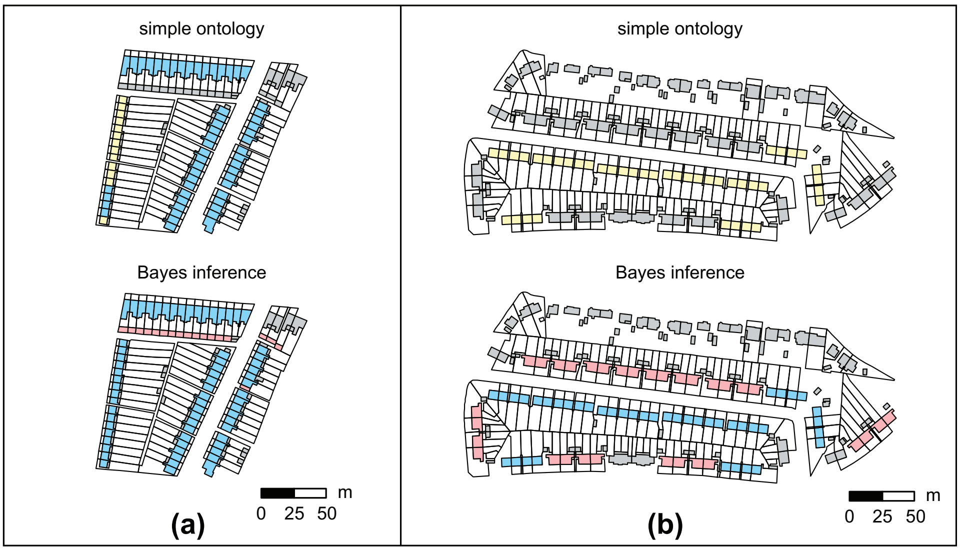


Fig. 9. Typical false positives in Bayesian inference. Colouring as in Fig. 6. OS MasterMap data Ordnance Survey ©Crown Copyright. All rights reserved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

8.2. Scalability considerations

The expenditure of time is more dependent on the low level algorithms involved than on the inference process itself, and therefore highly dependent on the actual ontology; we therefore constrain to the argument that the approach is practical. The inference of terraced houses took approximately 8 min for the Norwich study area (134,524 database objects) on a 2.66 GHz Xeon processor (single task). We therefore argue that the approach is practical, considering that it will typically be run as an off-line process for semantic enrichment of spatial databases rather than in real-time.

The necessity of defining training samples when joint probability distributions cannot be provided by a human operator can be seen as advantage and drawback at the same time. On the one hand, thresholds or membership functions such as when applying fuzzy set theory (Fisher et al., 2004; Ladner et al., 2003) do not have to be specified, but can be estimated from the training data. This is beneficial when knowledge about the domain is incomplete; for instance, clues in the literature about what 'small areas' means for terraced houses are rather vague. The downside is the effort that goes into the selection and tagging of training samples. When using kernel density for estimating probability distributions, density estimation effectively takes place in an *n*-dimensional feature space, which is created by the relations to sub-concepts. The more sub-concepts there are for a concept, the more training samples have to be defined to make sure that there are enough characteristic samples in each region of the feature space. This problem is known as the *curse of dimensionality* (Duda et al., 2001).

We also would like to comment on error propagation in the inference process. A concept definition usually relates to other concepts, whose instances are either asserted in the database, or have to be derived first. Poor accuracy in the derivation of related concepts leads to potential errors in the derivation of the composed concept. Since related concepts are derived independently, they should be checked for plausibility before continuing with inferring higher level instances. Therefore, the recognition process has to be supervised and is not fully automatic.

8.3. Benefits of the ontology-driven approach

The main benefits of this ontology-driven approach can be summarised as follows.

Enhanced transparency is provided since assumptions about the spatial structure of the geographic concepts are explicitly stated. Ontologies can be modelled and validated in collaboration with domain experts (making sure they are consistent with the experts' conceptualisation of reality), and different conceptualisations of the same terms can be compared, for example to reveal culturally different conceptions.

Enhanced flexibility is provided by being able to align the mapping of ontologies for different databases, or modify parts of an ontology to accommodate locally different settings.

Enhanced reusability is provided since it is a component-oriented approach that allows those parts that have to be implemented in spatial algorithms to be re-used in the derivation of different concepts. For this to happen, basic algorithmic components that provide spatial measures have to be identified and published. They serve as vocabulary that can be used for constructing ontologies. For instance, `presenceOf` was mapped to a density estimation, which constitutes an algorithmic component. The same component can be re-used to define a variety of patterns, such as the extents of urban areas and woods (Chaudhry & Mackaness, 2008; Mackaness, Perikleous, & Chaudhry, 2008). The concept `row of houses` can be re-used to define semi-detached houses (containing exactly two instances of `house` instead of at least three) or so-called perimeter block developments, which are an arrangement of rows of houses along the roads of a roughly square block. The concept `terraced house` can itself be used to derive even higher level concepts, such as `residential area`.

9. Conclusions and future research

Ontologies of the geographical reality are important because they provide a basis for abstraction of cartographically relevant patterns over large scale changes and for different usages. Hence the automated semantic annotation of spatial databases is a key success factor in support of automated map generalisation. In this paper, a framework for ontology-driven pattern recognition was presented. First, knowledge about the spatial structure of urban concepts is collected in an ontology. Then, the ontology is concretised by mapping it to measurable units. Finally, inference is carried out using Bayesian decision theory, whereas machine learning techniques can be used to learn concept characteristics from examples.

Besides clarifying the benefits of using ontologies in spatial database enrichment, our research has shown that Bayesian networks are a suitable method to integrate vague knowledge about conceptualisations in cartography and GIScience. We have also shown that logic reasoning techniques should best be combined with a set of general algorithmic components in order to achieve satisfying results.

Our future work will focus on the implementation of more concepts (e.g., other residential house types such as semi-detached and detached houses; on residential areas as an aggregation of residential house types) and a further formalisation of the pattern recognition vocabulary; on the evaluation of the choices of algorithms for basic concepts and their influence on extraction results; and on human subject experiments to study where and how people visually detect concepts such as terraces.

Acknowledgments

The research reported in this paper is part of the PhD project of the first author. Funding by the Swiss State Secretariat for Education and Research (SER) through COST Action C21 (project ORUS, Grant No. C05.0081) is gratefully acknowledged. The authors are grateful to the British Ordnance Survey for provision of MasterMap® data.

References

Agarwal, P. (2005). Ontological considerations in GIScience. *International Journal of Geographical Information Science, 19*(5), 501–536.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (Eds.), *The description logic handbook*. Cambridge: Cambridge University Press.

Bertin, J. (1967/1999). *Sémiologie graphique: Les diagrammes–Les réseaux–Les cartes*. Paris: Les réimpressions des Éditions de l'École des Hautes Études en Sciences Sociales [Original edition published 1967 by Gauthier-Villars, Paris].

Brassel, K. E., & Weibel, R. (1988). A review and framework of automated map generalization. *International Journal of Geographical Information Systems, 2*(3), 229–244.

Chaudhry, O. Z., & Mackaness, W. A. (2008). Automatic identification of urban settlement boundaries for multiple representation databases. *Computers, Environment and Urban Systems, 32*(2), 95–109.

Conzen, M. R. G. (1969). *Alnwick, Northumberland: A study in town-plan analysis*. London: Institute of British Geographers.

Costa, P. C. G. d., & Laskey, K. B. (2006). PR-OWL: A framework for probabilistic ontologies. In B. Bennet & C. Fellbaum (Eds.), *Formal ontology in information systems: Proceedings of the international conference on formal ontology in information systems (FOIS 2006)* (pp. 237–249). Amsterdam: IOS Press.

Ding, Z., Peng, Y., & Pan, R. (2006). BayesOWL: Uncertainty modeling in semantic web ontologies. In M. Zongmin (Ed.). *Soft computing in ontologies and semantic web. Studies in fuzziness and soft computing* (Vol. 204, pp. 3–29). Berlin/Heidelberg: Springer.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: John Wiley & Sons.

Fisher, P. (1999). Models of uncertainty in spatial data. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems and science* (pp. 191–205). New York: John Wiley & Sons.

Fisher, P., Wood, J., & Cheng, T. (2004). Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers, 29*(1), 106–128.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2003). *Ontological engineering: With examples from the areas of knowledge management, E-commerce and the semantic web*. Berlin: Springer.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199–220.

Grünreich, D. (1992). ATKIS – A topographic information system as a basis for GIS and digital cartography in Germany. In R. Vinken (Ed.). *From digital map series to geo-information systems, Geologisches Jahrbuch* (Vol. A 122, pp. 207–216). Hannover: Federal Institute of Geosciences and Resources.

Hart, G., Johnson, M., & Dolbear, C. (2008). Rabbit: Developing a controlled natural language for authoring ontologies. In S. Bechhofer, M. Hauswirth, J. Hoffmann, & M. Koubarakis (Eds.), *The semantic web: Research and applications. Lecture notes in computer science* (Vol. 5021, pp. 348–360). Berlin/Heidelberg: Springer.

Jena (2009). <http://jena.sourceforge.net/> Retrieved 25.03.09.

Jones, A. N., & Larkham, P. J. (1991). *Glossary of urban form. Historical geography research series no. 26*. London: Institute of British Geographers.

King, A. D. (1994). Terminologies and types: Making sense of some types of dwellings and cities. In K. A. Franck & L. H. Schneekloth (Eds.), *Ordering space – Types in architecture and design* (pp. 127–144). New York: Van Nostrand Reinhold.

Klien, E. (2007). A rule-based strategy for the semantic annotation of geodata. *Transactions in GIS, 11*(3), 437–452.

Ladner, R., Petry, F. E., & Cobb, M. A. (2003). Fuzzy set approaches to spatial data mining of association rules. *Transactions in GIS, 7*(1), 123–138.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2000). *Remote sensing and image interpretation* (4th ed.). New York: John Wiley & Sons.

Lüscher, P., Weibel, R., & Mackaness, W. (2008). Where is the terraced house? On the use of ontologies for recognition of urban concepts in cartographic databases. In A. Ruas & C. Gold (Eds.), *Headway in spatial data handling. Lecture notes in geoinformation and cartography* (pp. 449–466). Berlin/Heidelberg: Springer.

Mackaness, W. A., Perikleous, S., & Chaudhry, O. Z. (2008). Representing forested regions at small scales: Automatic derivation from the very large scale. *The Cartographic Journal, 45*(1), 6–17.

Marshall, S. (2005). *Streets and patterns*. Abingdon, UK: Spon Press.

Muthesius, S. (1982). *The English terraced house*. New Haven and London: Yale University Press.

Ordnance Survey Ontologies. (2008). <http://www.ordnancesurvey.co.uk/oswebsite/ontology/> Retrieved 31.10.08.

OWL (2008). <http://www.w3.org/2004/OWL/> Retrieved 19.05.08.

Rice, J. A. (1988). *Mathematical statistics and data analysis*. Pacific Grove: Wadsworth and Brooks.

Russel, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach*. Upper Saddle River: Prentice Hall.

Sen, S. (2008). Framework for probabilistic geospatial ontologies. *International Journal of Geographical Information Science, 22*(7), 825–846.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Simpson, J., & Weiner, E. (1989). *The Oxford English dictionary*. Oxford: Oxford University Press.

Steiniger, S. (2007). *Enabling pattern-aware automated map generalization*. Ph.D. thesis, University of Zurich.

SWRL (2009). <http://www.w3.org/Submission/SWRL/> Retrieved 19.03.09.

Thomson, M. K., & Béra, R. (2008). A methodology for inferring higher level semantic information from spatial databases. In D. Lambrick (Ed.), *Proceedings of the GIS research UK 16th annual conference (GISRUK 2008)* (pp. 268–274). <http://www.unigis.org/gisruk_2008/proceedings.htm> Retrieved 31.03.09.

Yu, L. (2007). *Introduction to the semantic web and semantic web services*. Boca Raton, Florida: Taylor & Francis Group.

Zhang, X., Stoter, J., Ai, T. (2008). Formalization and automatic interpretation of map requirements. In 11th ICA workshop on generalisation and multiple representation, Montpellier, France. <http://ica.ign.fr/montpellier2008/program.php> Retrieved 31.03.09.

Zheng, H.-T., Kang, B.-Y., & Kim, H.-G. (2007). An ontology-based Bayesian network approach for representing uncertainty in clinical practice guidelines. In P. C. G. da Costa et al. (Eds.), *Uncertainty reasoning for the semantic web I. Lecture notes in artificial intelligence* (Vol. 5327, pp. 161–173). Berlin/Heidelberg: Springer.

Publication IV

Lüscher, P., & Weibel, R. (submitted 2010). Exploiting empirical knowledge for delineation of city centres from large-scale topographic databases. *Computers, Environment and Urban Systems*, revised manuscript submitted June 2011.

Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases

Patrick Lüscher and Robert Weibel

University of Zurich, Department of Geography

CH-8057 Zurich, Switzerland

E-mail: {patrick.luescher, robert.weibel}@geo.uzh.ch

Keywords: Pattern recognition, topographic database, semantic modelling, urban structure, city centre

Abstract

Current topographic databases rarely represent higher order geographic phenomena, such as city centres. However, such concepts are often referred to by humans and used in various forms of spatial analysis. Hence, the value and usability of topographic databases can greatly be improved by methods that automatically create such higher order phenomena through cartographic pattern recognition techniques, departing from the very detailed, geometry-oriented representations of topographic databases. As many higher order phenomena are only vaguely defined, this paper develops and evaluates a methodology to acquire definitional knowledge about geographic phenomena by participant experiments and use this knowledge to drive the cartographic pattern recognition process. The method is applied to acquire knowledge about British city centres and delineate referents of city centre from topographic data. City centres produced for ten British cities are compared to areas derived from alternative sources. F_1-scores between 0.45 and 0.88 are achieved, suggesting that the delineation produced plausible city centre areas. The benefits of our work are better (and user-driven) descriptions of complex geographic phenomena that can form the basis for accurately enriching topographic databases with additional semantics, thus yielding added value for the data producer and the end user.

1 Introduction

National Mapping Agencies (NMAs) and other data producers maintain and disseminate topographic datasets at the very fine scale. Being designed as general purpose products, these datasets offer a wealth of (mainly geometric) information about individual objects. However, they do not the model higher order geographic phenomena required by many applications. For example, they model buildings and parking spaces, but not hospital complexes, districts and settlements (Chaudhry & Mackaness, 2008a; Chaudhry, Mackaness, & Regnauld 2009; Lüscher, Weibel, & Burghardt, 2009); they model height fields, but not the extent of hills, valleys and mountain ranges (Chaudhry & Mackaness, 2008b; Straumann, 2010).

Improving their datasets by providing more of such higher level semantics could help NMAs and other data producers to establish a more user-driven access to geographic information (Hart & Greenwood, 2003; Davies, Wood, & Fountain, 2005). It allows better adapting representations to the way how people conceptualise geographic space. The study of the body of knowledge that people have about the surrounding geographic world is termed naïve geography (Egenhofer & Mark, 1995). Human spatial reasoning is chiefly qualitative, i.e. based on spatial relations and regions (Egenhofer & Mark, 1995; Montello, 2003). Representing geographic regions is thus beneficiary for many applications such as geographic information retrieval, navigation, and building gazetteers (Heinzle, Kopczynski, & Sester, 2003; Purves et al., 2007, Montello, 2003). For example, people might be interested in answers to queries such as “Where are city centre hotels?” Furthermore, having higher order phenomena in the database allows NMAs to respond better to customer requirements. Professionals of various disciplines maintain that concepts related to urban area and place, such as settlement, neighbourhood, townscape, and urban structure, are key spatial concepts (Davies, Holt, Green, Harding, & Diamond, 2009). This is often reflected in medium scale maps and maps for urban planning which emphasise urban structure (Steiniger, Lange, Burghardt, & Weibel, 2008).

The map generalisation community is engaged in creating such higher order representations from topographic datasets through model generalisation techniques. *Model generalisation* includes various operations to abstract, aggregate, re-classify and reduce representations in a topographic database, without aiming at visual presentation; processes envisaged to optimise visualisation quality are subsumed as *cartographic generalisation* (Weibel, 1997). A number of model generalisation techniques have been proposed based on taxonomies. Their application is however restricted to small changes in representation (Chaudhry, 2008a). Achieving more drastic abstractions requires that the semantics of the phenomena is modelled in a prototypical sense (Nyerges, 1991; Mackaness, 2006). Humans seem to define categories in terms of prototypes that contain the most representative attributes within that category (Rosch, 1978; Mennis, Peuquet, & Qian, 2000). Categories have a graded internal structure, i.e. some objects are more typical instances of a category than others.

This paper presents a study and methodology to define and delineate UK city centres from topographic data. The city centre, described as the "heart of the city" by Murphy and Vance (1954), is of particular interest due to its function as nucleus of both business and community activities within the city. Dramatic changes of city centre economy in recent decades raised issues of vitality and sustainability. In the context of city centre regeneration, numerous studies investigated topics such as retail development (e.g. Lowe, 2005; Thurstain-Goodwin & Unwin, 2000), visitor activity patterns (Bromley, Tallon, & Thomas, 2003), community safety (Townshead & Pain, 2000), and city centre access and pedestrian movement (Borgers & Timmermans, 1986).

Prototypical definitions are challenging to acquire for phenomena that are only vaguely defined, such as a city centre. The key aim of this paper is to establish a user-driven methodology to capture models of higher order geographic phenomena for model generalisation. We conduct an online participant experiment to capture the prototypical meaning of a city centre, present a procedure to delineate city centres from topographic data, and finally evaluate the model and the delineation procedure by comparison of the results to other sources. The present paper extends on research developed in previous papers (Lüscher, Weibel, & Mackaness, 2008; Lüscher et al., 2009), all pursuing the objective of enriching common, cartography-oriented spatial databases with high level semantics. The research questions we aim to address in this paper are as follows:

1. In general, how can empirical knowledge be formalised to delineate higher order phenomena from topographic databases?
2. More specifically, what are the defining elements of a city centre?
3. How can the produced regions be evaluated?

The remainder of this paper is organised as follows. After a review of related work in Section 2, the methodology of eliciting the city centre prototype and the computational procedure to delineate city centres from topographic data are introduced in Section 3. This is followed by the presentation and evaluation of the results in Section 4. Section 5 provides a discussion and Section 6 then concludes the paper with an outlook.

2 Related work

City centres: An early method to delineate central business districts was proposed by Murphy and Vance (1954). For each urban block, the amount of floor space devoted to retailing and commercial activities were used to compute indices of central business activity. A similar approach to delineate town centres was presented by Thurstain-Goodwin and Unwin (2000), aiming at monitoring of urban retail activities. Employment and floor space data was used to create continuous surfaces of town-centric activity. Montello, Goodchild, Gottsegen and Fohl (2003) conducted experiments in delineating 'downtown' by asking people in the street to draw an outline on a paper map. More recently, crowd-sourcing methods were investigated to delineate vernacular areas. Hollenstein and Purves (2010) used georeferenced images from flickr.com to investigate the vernacular use of city core terms.

Pattern recognition from topographic data: Specialised techniques exist for the recognition of urban structures and patterns, using geometric algorithms and/or statistical methods. Many of these techniques focus on the key feature classes defining the urban environment, roads (e.g. Heinzle & Anders, 2007) and buildings (e.g. Regnauld, 2001), and were originally devised to optimise cartographic quality in map generalisation.

A number of techniques were elaborated to abstract topographic datasets to higher order representations. Most of these make use of morphological variables only. For example, approaches exist to delineate settlement boundaries, based on building size and density (Joubran & Gabay, 2000; Boffet, 2001; Chaudhry and Mackaness, 2008a). Graph-based measures and building morphology were also used to separate areas of urban land use and period of construction (e.g. Barr, Barnsley & Steel, 2004; Steiniger et al., 2008). Boffet (2001) aggregated urban blocks into districts by means of land use and morphology. She also proposed the use of building density as a means to isolate city centres. However, she did not attempt a systematic study.

There have been proposals in the literature to explicitly model geographic phenomena to improve transparency and expressiveness of the model generalisation process. This means to model the semantics of geographic phenomena as sets of properties and (spatial) relations to other concepts. Mallenby (2007) used such an approach for detecting water features. Thomson (2009) presented a method to separate knowledge from pattern recognition algorithms by ontological reasoning on building types and land use categories. Previous work involving the authors has also successfully exploited the use of ontologies in detecting urban house types (Lüscher et al., 2008), including reasoning in the presence of vagueness (Lüscher et al., 2009).

Acquiring and modelling geographic phenomena is challenging, particularly if the phenomenon is only vaguely defined. Conceptualising geographic phenomena as they are understood and used by people, however, would make the derived representations more useful for many applications as discussed in Section 1. Hence, this paper explores the use of participant experiments to capture semantics and subsequently formalises this empirical knowledge for model generalisation. A second aim of the paper is to develop a procedure to spatially delineate city centres from topographic databases.

3 A method to delineate city centres

3.1 Overview

Figure 1 shows an overview of the proposed procedure of computing and evaluating a city centre. The datasets used for the experiment are explained in Section 3.2. To gain a solid basis for the physical and functional characteristics that constitute a British city centre according to a broad group of people, a participant experiment was carried out (Section 3.3). Based on the analysis of the participant experiment a model of city centre typicality was established which was used to compute city centre typicality values at each point of a regular raster (Section 3.4). The city centre model consists of (groups of) features that are typical or untypical of a city centre and hence have a positive or negative influence on perceived city centre typicality. For each group of features a separate individual city centre typicality surface is computed. The individual typicality surfaces are finally combined to a single city centre typicality value by weighted summation. A crisp city centre area was obtained by applying region growing and a threshold to the continuous city centre typicality surface (Section 3.5). Finally, we suggest several ways to evaluate the plausibility of the computed city centre areas (Section 3.6).

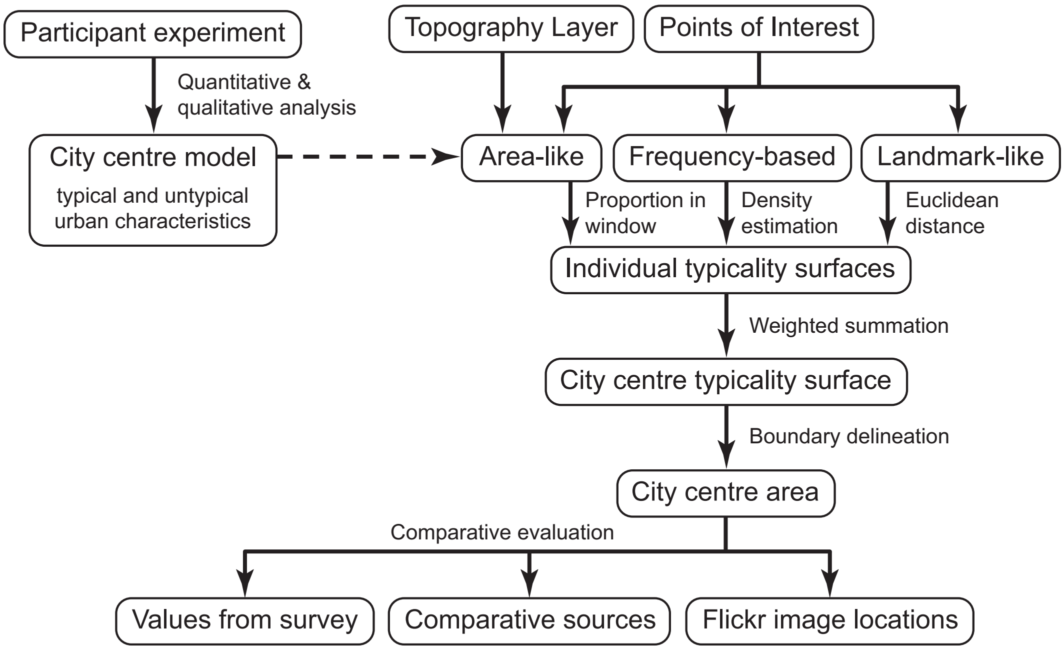


Fig. 1. Overview of procedure for computing a city centre.

3.2 Datasets

The following datasets provided by the Ordnance Survey of Great Britain (OS) were used. All are vector datasets.

OS MasterMap® Topography Layer roughly represents what can be seen on a topographic map. The granularity corresponds to a scale of 1:1,250 in urban areas. The polygon features in the Topography Layer form an exhaustive representation of land cover. The primary classification of the features is provided by an attribute that assigns each feature one of currently 21 groups, such as 'Building', 'Natural Environment', and 'Road or Track'.

OS Points of Interest (POI) is the main dataset used for obtaining functional information. It covers commercial addresses and features of interest classified into a three-level taxonomy. The topmost level encompasses 9 classes: Accomodation; Eating and drinking; Attractions; Commercial services; Sport and entertainment; Education and health; Public infrastructure; Retail; Manufacturing and production; and Transport. The most detailed level contains more than 600 classes.

OS MasterMap® Address Layer 2 is a second Points of Interest dataset offered by the Ordnance Survey. In comparison to the OS Points of Interest dataset, it additionally encompasses residential addresses. However, our evaluation revealed that the coverage of commercial establishments is rather bad. Hence, both datasets were fusioned into a single functional features dataset, taking residential features from Address Layer 2, and all other features from Points of Interest.

OS Strategi® was derived by digitising Ordnance Survey's 1:250 000 scale maps. In our experiments, the settlement extents of OS Strategi® were used to delimit the study areas.

The experiments were carried out for the following 10 British cities: Birmingham, Bristol, Cardiff, Leeds, Liverpool, Glasgow, Manchester, Nottingham, Sheffield, and York. The selected cities range in population size from 198,800 (York) to 1,028,700 (Birmingham) (Office for National Statistics, 2010), and provide a variety of different topographical settings and urban history.

3.3 Participant experiment

An experiment in the form of an online questionnaire was developed to elicit a prototypical model of a city centre from a broad range of people. The results of the questionnaire were then used to build a model of city centre typicality (or 'city centreness', Section 3.4). Additionally, a part of the questionnaire was used to verify the model output (Sections 3.6.3 and 4.3).

3.3.1 Design and procedure

The questionnaire was implemented as a set of HTML forms. It was organised into three parts which the participants had to answer in a fixed order. The full questionnaire is provided as electronic supplementary material and can be downloaded from the journal's website. In the following, the two relevant experiments for defining properties of a city centre are presented and discussed.

The first part of the questionnaire was meant to capture an uninfluenced, individual image of a city centre. It contained experiments where participants had to describe separately frequent activities, important facilities and services, and optionally physical characteristics of a city centre. Answers were to be provided as free text. The task was introduced as follows:

> Please define, briefly, in what aspects a city centre differs from other areas of a city.

To render the task more concrete, we asked specifically for services and facilities:

> Please indicate: What kind of services & facilities do you expect to find there (in comparison to other areas)?

In the second experiment, the participants were presented a list of urban features and asked to decide whether the features were typical of a city centre. The list is a subset of the full OS

Points of Interest taxonomy which was compiled by considering experiences made in previous studies on city centre use (e.g. Bromley et al., 2003; Tallon & Bromley, 2004) and features visible on common topographic maps. Answers were possible on an ordinal scale between -2 (very untypical) and +2 (very typical). The instructions for grading typicality read as follows:

The following lists contain certain types of concepts that are to be found commonly in urban areas. Please indicate the degree to which they are typical for a city centre.

Select **'Very typical'** if:
- You think that the concept is typically only found within a city centre.
- If you think the best location to find many of the concepts is a city centre.
- If you think the concept is very characteristic for a city centre.

Select **'Very untypical'** if you wouldn't expect such a concept in a city centre.

Select **'Can be either'** if you think the concept can be found commonly within a city centre as well as outside of it.

If you are not sure about the meaning of a concept and can't answer a question, select **'Don't know'**.

3.3.2 Participants

Participants were recruited in two ways. Firstly, an invitation email was sent to several British academics for distribution among their peers and students. Secondly, the link was published in the bulletin boards of two websites that focus on urban planning and geography (www.skyscrapercity.com and www.geograph.org.uk). To provide an incentive, three book vouchers of £50 each were drawn amongst all participants. In the course of a month (March 2010), 101 completed and valid questionnaires were obtained this way.

70.3% of the respondents were male. Similarly, the age of the respondents is biased towards younger people (Figure 2).

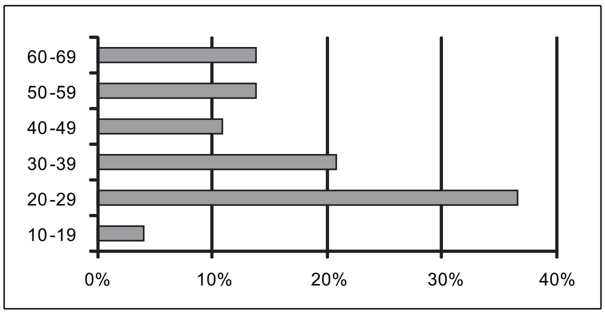


Fig. 2. Age structure of respondents.

Participants were also asked to indicate current and former places of residence. 40.6% of the respondents always lived at the same place; 36.6% moved, but always within the UK. 82.2% had been living in the UK for longer than 10 years. The geographical distribution of the respondents (Figure 3) shows peaks where the participating academic institutions are located, but the respondents are reasonably well scattered across the United Kingdom. 14.9% of the places of residence are rural areas, and 85.1% urban areas. 70.3% of the respondents indicated a place of residence that has city status.

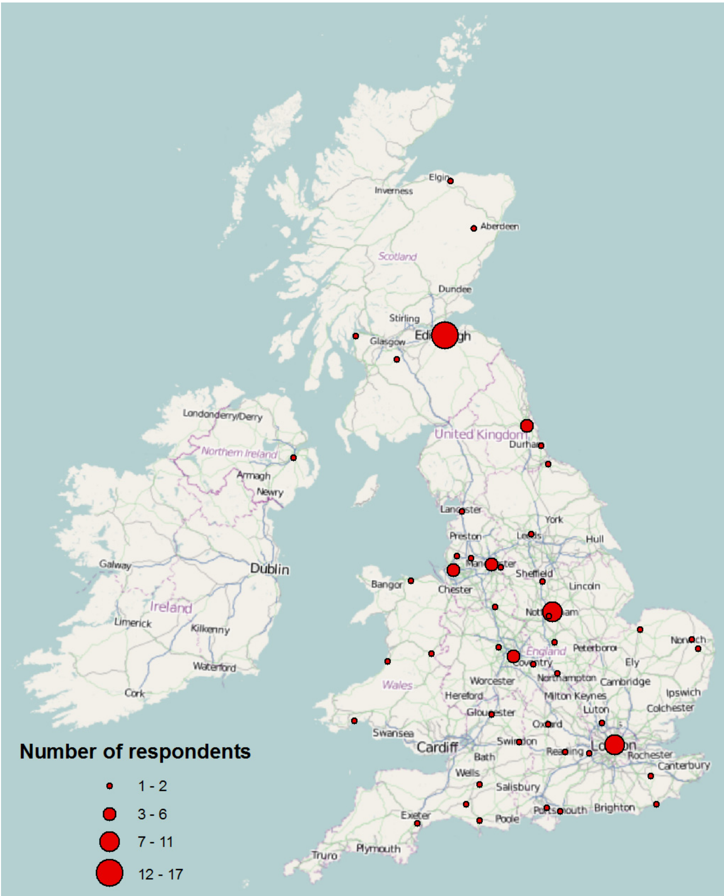


Fig. 3. Places of residence of respondents. Background mapping © OpenStreetMap contributors, CC-BY-SA.

3.3.3 Analysis of participant experiment

Figure 4 shows the results of the urban feature grading experiment. The numbers in brackets are cross-references to equivalent concepts in Table 1.

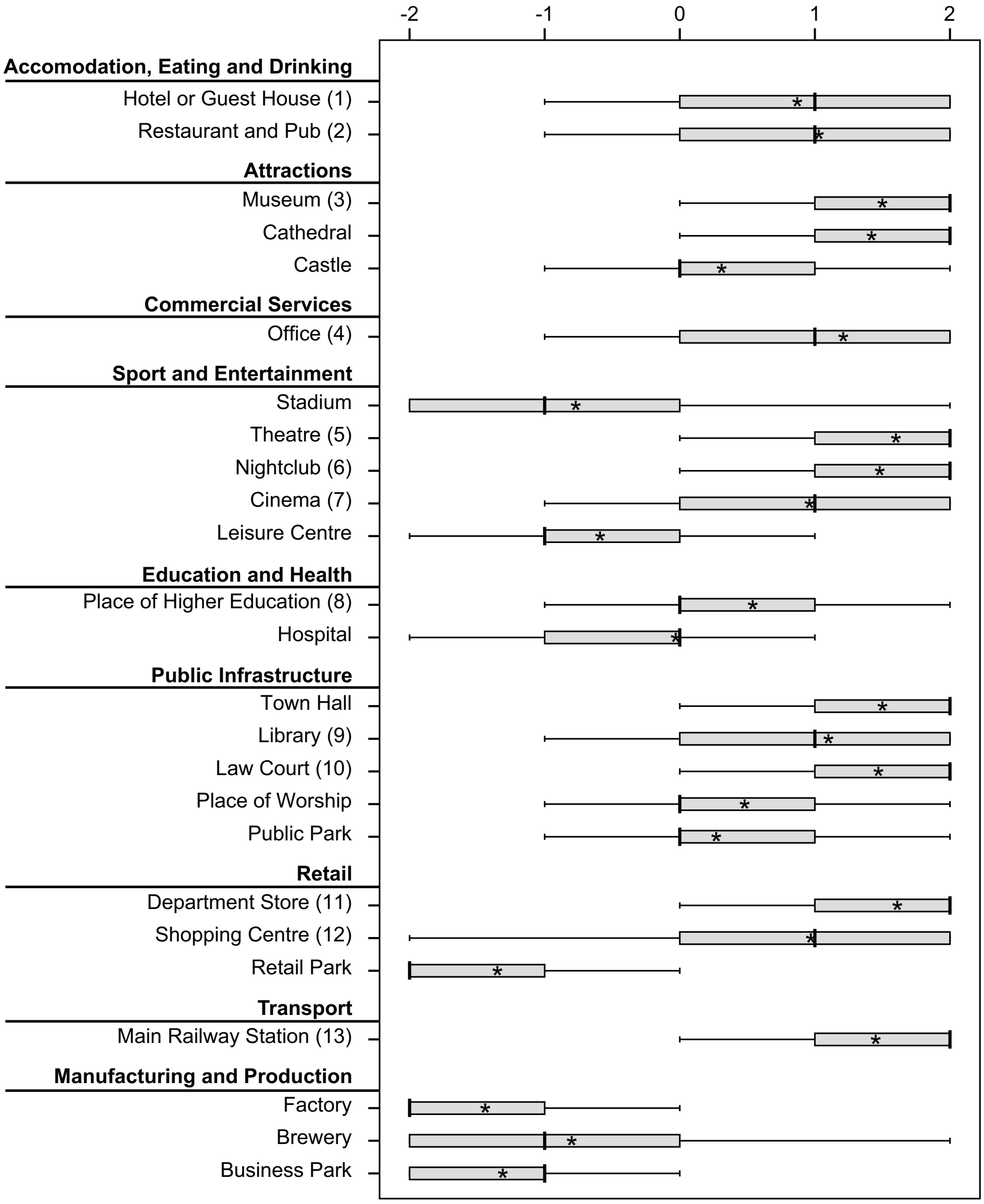


Fig. 4. Typical (+) and untypical (-) concepts for city centres. The box plots indicate mean (*), median (thick line), and 1st and 3rd quartile (width of boxes). Whiskers include approximately 95% of the responses.

The task to enumerate important services and facilities in city centres resulted in lists of items by each participant, such as *"Shops, Restaurants, Chain Bars, Shopping Centres"* and *"Cafes and restaurants, Shops, Lots of bus stops, Railway stations"*. Often the participants qualified the items they named. Some respondents wrote for example *"more specialised shops"*, *"more diverse restaurants"*, or *"denser/richer variety of shops"*. However such qualifiers were not used. Rather, the occurrence of each concept (such as "bar" and "restaurant") was counted. There were in total 50 different concepts named by the participants. Table 1 shows the concepts named by at least 5% of respondents.

Type of facility	No. of mentions in % of respondents
Accommodation, eating and drinking	
Hotel (1)	4.95
Restaurant (2)	42.57
Pub (2)	24.75
Café	9.91
Attractions	
Museum (3)	21.78
Art Gallery	12.87
Commercial services	
Office (4)	11.88
Sport and entertainment	
Theatre (5)	22.77
Bar	17.82
Night club (6)	12.87
Cinema (7)	10.89
Concert hall / venue	5.94
Education and health	
University (8)	6.93
Public infrastructure	
Civic services & seats of parliament (~9)	20.79
(Main) Library (10)	9.90
Retail	
Shops (boutiques & special goods)	67.33
Department store (11)	8.91
Shopping centre (12)	6.93
Bank	13.86
Manufacturing and production	
None named	
Transport	
Transport hubs (Railway & coach terminals) (~13)	37.62
Dense public transport	21.78

Table 1. Typical facilities named by the participants.

3.4 Operationalisation of city centre typicality

3.4.1 City centre typicality surfaces

The two tasks in the questionnaire were analysed in combination to obtain a model of perceived city centre typicality (or 'city centreness'). According to the questionnaire, city centre typicality is high if there is a high concentration of places for eating out and for shopping for special goods, as these concepts were mentioned frequently by respondents. Restaurants received rather low typicality values (Figure 4). This is explained because they occur also outside city centre, but in lower concentrations. Concepts like transport hubs, town halls, and cathedrals occur only once (or few times) in a city, but create a zone of high city centre typicality. High city centre typicality is also produced by the absence of business parks and manufacturing, while features such as castles or hospitals do not influence city centre typicality. The survey was analysed in this way to compose groups of features that influence city centre typicality in a positive or negative way, respectively. The final list of characteristics is shown in Table 2.

For each of the items in Table 2, a separate typicality surface was computed (details follow in Sections 3.4.2 to 3.4.4). The individual surfaces were finally aggregated into a city centre typicality surface by weighted summation (Equation 1).

$$typicality_{citycentre} = \frac{\sum_i w_i \times typicality_i}{\sum_i w_i} \quad (1)$$

The weights w_i of the individual typicality surfaces $typicality_i$ were determined by considering city centre typicality of urban features indicated by the participants in Figure 4 and Table 1. For example, theatres and museums were named frequently and indicated as very typical since they are hardly located outside of city centres. Thus, they were assigned a weight of 1. Office-based services were indicated as somewhat typical and thus received a weight of 0.5. It was also observed in the experiments that industrial and suburban residential areas (i.e. terraced, detached and semi-detached housing) are seen as very untypical for city centres and indeed they often serve as bounding features for a city centre. The high negative weight of -4 assigned to these features cancels out effects of nearby city centre features, such that raster cells within industrial and residential areas always have low city centre typicality values. A similar, but less strong negative influence was observed for the amount of open ground.

Typicality surface	Type	Weight
Accommodation, eating and drinking		
Places to eat and drink (restaurants, pubs, etc.)	F	0.75
Attractions		
Museums and art galleries	F	1
Cathedrals	L	0.5
Commercial services		
Office-based services (stock trading, architects, etc.)	F	0.5
Sport and Entertainment		
Night clubs, amusement arcades	F	1
Theatres, concert halls	F	1
Public infrastructure		
Civic services (consular services, courts, etc.)	F	1
Town hall	L	0.5
Main libraries	L	0.125
Retail		
Boutiques and special goods shops, department stores	F	1
Banks and retail services	F	0.25
Retail parks	F	-1
Transport		
Public transport hubs (main railway stations, coach stations)	L	1
Public transport services (bus stations, tram stations, etc.)	F	0.75
Manufacturing and Production		
Industrial areas	A	-4
Suburban Features		
Suburban residential areas	A	-4
Natural open ground (groves, pastures, bodies of water)	A	-2

Table 2. Individual typicality surfaces. Types: F = Frequency-based, L = Landmark-like, A = Area-like.

From the analysis of the participant experiment it became clear that features influence city centre typicality in three different ways. Firstly, features such as shops, retail services, and bus stops characterise city centres by their concentration (and sometimes diversity). Hence, a frequency-based typicality surface is estimated by Kernel Density Estimation (KDE, Section 3.4.2). Second, certain features (e.g. town halls and railway terminals) occur only once (or few times) in a city, but are nevertheless important features in structuring the urban landscape; hence they are termed 'landmark-like'. Rather than the density, the distance to such features is relevant (Section 3.4.3). Thirdly, large urban regions such as residential districts and industry parks cannot be modelled by points alone. Industrial areas, for example, are comprised of many features, such as factories, office buildings, and open surfaces, whereas the POI dataset generally only covers the locations of head offices. Thus, such areas have to be created first by means of specific algorithms. Their influence is measured by their proportion in a circular window around each raster pixel (Section 3.4.4). The creation of typicality surfaces for each of the three categories is now described.

3.4.2 Modelling of frequency-based characteristics

For individual establishments, a surface was computed using Kernel Density Estimation (KDE). KDE requires two parameters: The bandwidth and the kernel function, which determines the weighting of the points. In our case, we used a quadratic kernel function. While it is reported that the choice of the kernel function has little influence on the results (Lloyd, 2007, p. 184), the selection of bandwidth is more important. A number of data-driven methods exist to estimate bandwidths objectively (Jones, Marron, & Sheather, 1996). A plug-in bandwidth estimator provided by Duong (2007) was employed to sample bandwidths for a subset of the typicality surfaces. Based on the estimates, it was decided to use a single bandwidth of 350 m for all surfaces to simplify matters and improve comparability. Each surface was subsequently normalised, such that 0 = *minimum typicality* within the study area, and 1 = *maximum typicality*.

3.4.3 Modelling of landmark-like features

For each of the landmark concepts a typicality surface was computed as a function of the Euclidean distance to the landmark feature. Landmarks can be considered as anchors of cognitive representations of space (Winter, Tomko, Elias, & Sester, 2000). Landmarks can be differentiated based on prominence, uniqueness, and salience. Global landmarks, such as the landmarks in this study, are used for referencing from larger distances in a city. The normalisation for landmark typicality surfaces thus assumes a maximum distance of 3 km, corresponding to the size of a large city centre (e.g. Liverpool). Cells further away than 3 km receive a typicality of 0, and distances between 0 km and 3 km are linearly scaled to values between 1 and 0.

3.4.4 Modelling of area-like characteristics

While natural open ground is coded in the Topography Layer (in the form of natural areas and water), residential neighbourhoods and industrial areas are themselves complex concepts that were derived in a separate procedure beforehand. An approach for reliably extracting suburban residential buildings from topographic data was shown in a previous publication (Lüscher et al., 2009). Chaudhry et al. (2009) presented an approach to extract functional sites (such as airports and hospitals) from topographic data. The approach used here follows the idea of Chaudhry et al. (2009), but in a simplified form as there is no iterative growing involved. The functional features were intersected with buildings from the Topography Layer to enrich buildings with functions. Then, the algorithm proceeded as described in Table 3.

Residential areas	Industrial areas
Extract residential-only buildings	Extract all buildings that have an industrial function, whereas business services are also allowed
Extract yards that touch the residential buildings	Extract open, manmade and natural surfaces that touch the industrial buildings
Merge residential buildings and yards and dissolve to preliminary residential areas	Merge industrial buildings and open surfaces to preliminary industrial areas
Keep only residential areas that have at least 5 residential buildings	Keep only industrial areas where the portion of industrial building area exceeds 50 % of the total building area and that have a total area > 1000 m^2

Table 3. Steps for delineating residential and industrial areas.

Figure 5 illustrates residential areas obtained in this way. Areas of terraced and semi-detached housing are delineated as suburban residential, while the high street area in the centre and the park in the eastern part of the extract are excluded.

A typicality surface for each type of urban district was obtained by computing the portion of the respective land use within a circular window of 250 m size. The window size is different to the one used for KDE because all features within the window have a constant weight, while the quadratic kernel weights distant points less than points near the window centre. The window sizes were thus chosen such that the volumes enclosed by the windows are approximately equal.

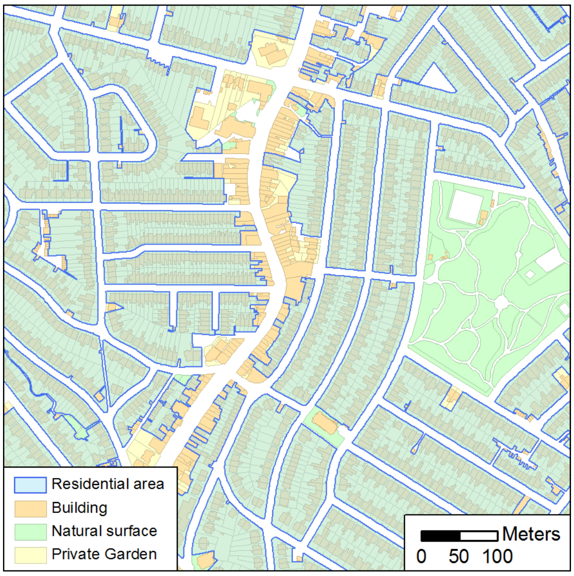


Fig. 5. Extracted suburban residential areas. (Ordnance Survey© Crown Copyright. All rights reserved).

3.5 Boundary formation

While it is possible to produce a fuzzy city centre region from the typicality surface, it makes more sense to produce crisp boundaries for many applications, such as cartographic visualisation, query processing, and urban planning (Couclelis, 1996). A region growing algorithm was developed for automatically determining the boundaries of a city centre. The algorithm initialises a city centre area with the cell of highest city centre typicality within a study area. The area is then iteratively enlarged by adding the cell of highest typicality among all cells that are adjacent to the current area. The process stops when the collected area reaches a certain threshold. The obtained city centre boundaries are finally generalised by morphological operations (i.e., erosion and dilation of the polygon) (Millward, 2004).

The most critical part of the process is finding an appropriate threshold for stopping the growing process. In our case, a best-fit value of 0.5 was chosen by considering comparative city centres (cf. Section 3.6.1). Figure 6a shows the evolution of city centre typicality during the growing process. A city centre is delineated when computed city centre typicality drops below 0.5, i.e. its typicality line enters the grey shaded area in Figure 6a. Spikes of increased typicality occur when the growing process captures secondary areas of high centre typicality. Furthermore, it can be seen that the typicality behaves very similarly in all cities and decreases in a power or logarithmic function with increasing area and hence with increasing distance to the point of highest typicality. Figure 6b shows the progression of the algorithm in Bristol.

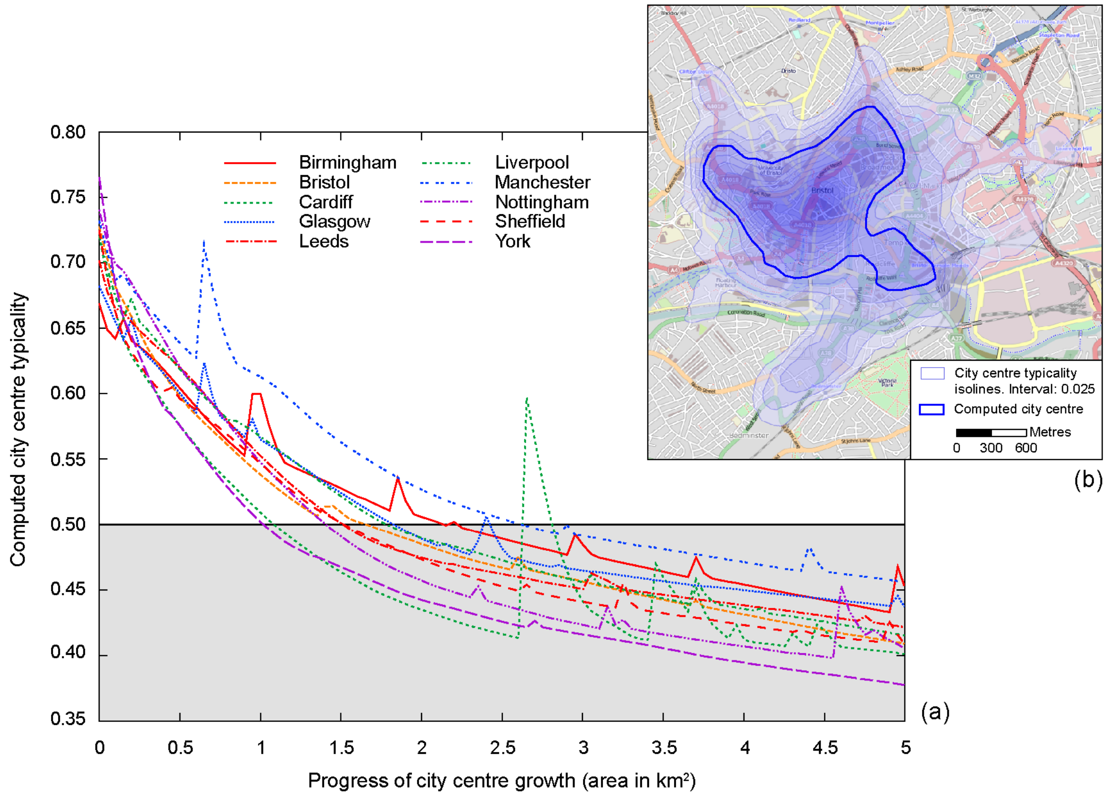


Fig. 6. (a) Plot of city centre typicality against increasing area and (b) contour map of city centre typicality in Bristol. Background mapping © OpenStreetMap contributors, CC-BY-SA.

3.6 Comparative evaluation

It is obviously challenging to evaluate vague geographic phenomena such as city centres as there cannot be definite reference data. Here, we propose three different methods for assessing produced city centre boundaries.

3.6.1 Comparative city centre representations

The research studied alternative sources for delineating city centre areas and using them as comparative representations to validate the boundaries produced by our approach.

The web was manually searched for representations of the city centre of each city. The search mainly focused on maps which explicitly designated a city centre area, such as tourist maps, or bus maps. Furthermore, Wikipedia provides narrative descriptions of the extents of some city centres. These descriptions were interpreted and mapped. For each city, we created between one and four alternative representations in this way. Collectively, these descriptions give us hints about the extent of the *vernacular* city centre, but we prefer to call them *comparative city centres* rather than *reference city centres*, as they are themselves vague interpretations, represent an individual opinion, or are the result of a political compromise and are therefore different from people's conceptualisation of a city centre. For example, city centre designations on tourist maps may be biased due to the focus on sites of interest for visitors, i.e., sites of historic or cultural significance. Narrative descriptions on Wikipedia such as *"bounded north by St Pauls and Easton, east by Temple Meads and Redcliffe, and west by Clifton and Canon's Marsh"* (http://en.wikipedia.org/wiki/Bristol_city_centre, accessed 14.04.2010) are difficult to confine and might even contain contradictory statements.

The number of representations obtained depends on the number of sources found and their agreement. For example, sources of comparative centres for Glasgow all agree on the extent of the city centre; hence there is only one comparative representation. There is more disagreement for Bristol, where four different interpretations of the city centre extent were acquired.

3.6.2 Volunteered geographic information

As discussed in Section 2, information from the internet can be used as a proxy of people's vernacular geographic knowledge. The procedure used in this work follows Hollenstein and Purves (2010) who used flickr.com as source of information. Flickr.com is a website where people upload images and describe them by means of tags. It is also possible to attach a geographic location to the image. Flickr provides a web API for automatically searching and downloading such information.

Locations of georeferenced images tagged as 'city centre' were downloaded from flickr.com. For each study area, a distribution of image locations was obtained in this way. Relatively few image locations were available for many cities, such that no representative pattern could be deduced. The comparison hence focuses on four cities: Birmingham, for which 213 locations contributed by 58 people were available; Glasgow (325 locations contributed by 61 people); Liverpool (248 locations contributed by 39 people); and Manchester (421 locations contributed by 90 people).

Vague footprints were created from the point distributions by means of kernel density estimation (KDE) as described by Hollenstein and Purves (2010). The area within the 80% volume contour was selected for quantitative evaluation as it seemed to produce the most plausible city centre areas in the four cities.

3.6.3 Rating of panoramic images

A task of the participant experiment consisted of a series of 360° panoramic images showing urban scenes (Figure 7). In total 15 panorama sites were prepared, out of which a respondent had to judge 10 randomly selected sites. The sites were selected to cover a range of different categories of environment. 12 of the images were located in Bristol; additional 3 images were selected from Manchester to provide a more varied coverage of city centre situations. The panorama sites also showed rather prototypical vistas. In particular, we avoided situations such as through roads bordered by shops, or streets that are within the city centre, but that are poor on features indicative of a city centre. Such situations are difficult to judge from the images alone.

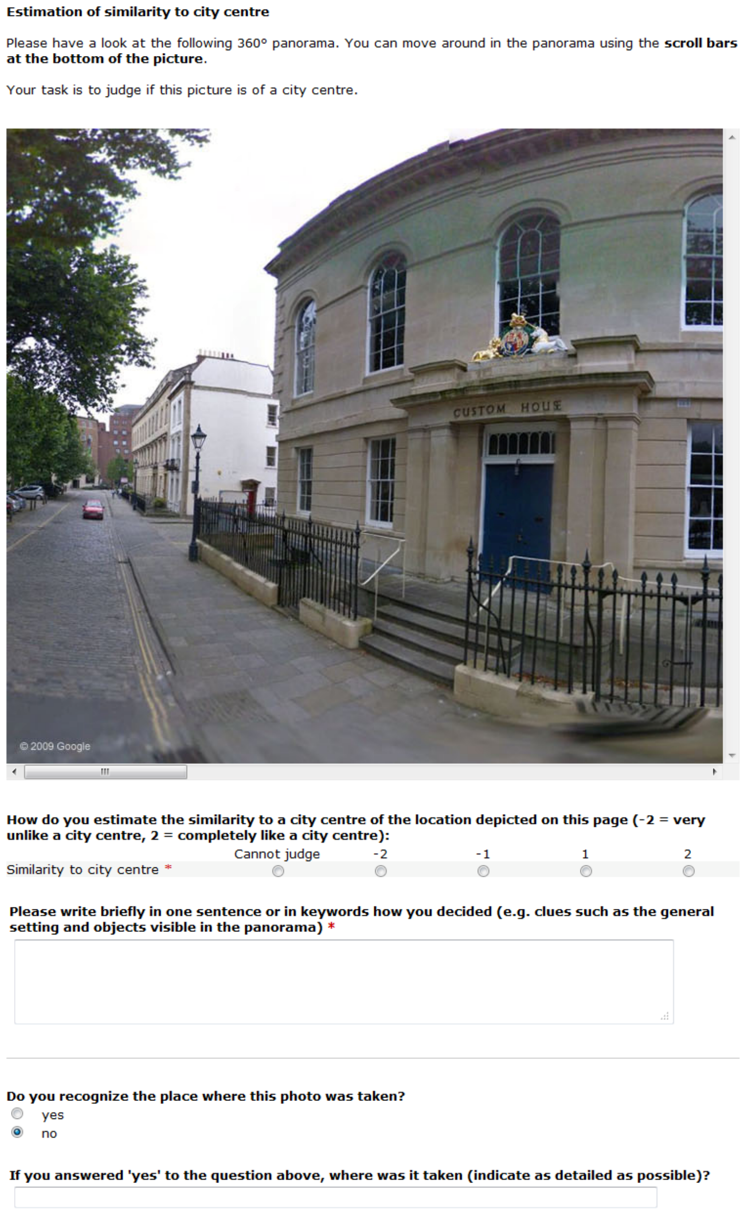


Figure 7. An example stimulus for estimating city centre typicality based on 360° panoramic images.

The participants had to decide on the degree to which the scene conformed to a city centre. We also asked whether the participants recognised the place shown on the image, and if so, to indicate its location. However, only one site (Spring Gardens in Manchester) was frequently recognised (details can be found in the electronic supplementary material). In Section 4.3, the typicality values estimated by the participants are compared to the computed typicality values.

Links to the 15 panorama sites are included in the electronic supplementary material to this article.

4 Results

4.1 Computed city centre boundaries

Figures 8 and 9 show the computed city centres versus the comparative city centres for each city. Table 4 makes a quantitative comparison of the overlap between computed and alternative city centre areas. It shows precision and recall values, and the F_1-score, which is the harmonic mean of precision and recall. In Equations 2 and 3, $a_{computed}$ denotes the city centre area as delimited by the algorithm, $a_{comparative}$ denotes the area of comparative/Flickr city centre representations, and $a_{overlap}$ denotes the area where computed and comparative/Flickr city centres overlap.

$$precision = \frac{a_{overlap}}{a_{computed}} \tag{2}$$

$$recall = \frac{a_{overlap}}{a_{comparative}} \tag{3}$$

$$F_1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

If there are multiple comparative areas for a city, the comparison is twofold: against the intersection of the comparative areas, which act as a narrow interpretation of the city centre, and against the union of the comparative areas as a loose interpretation of a city centre. Values for intersection and union are equal where there is only one comparative city centre (i.e. Glasgow).

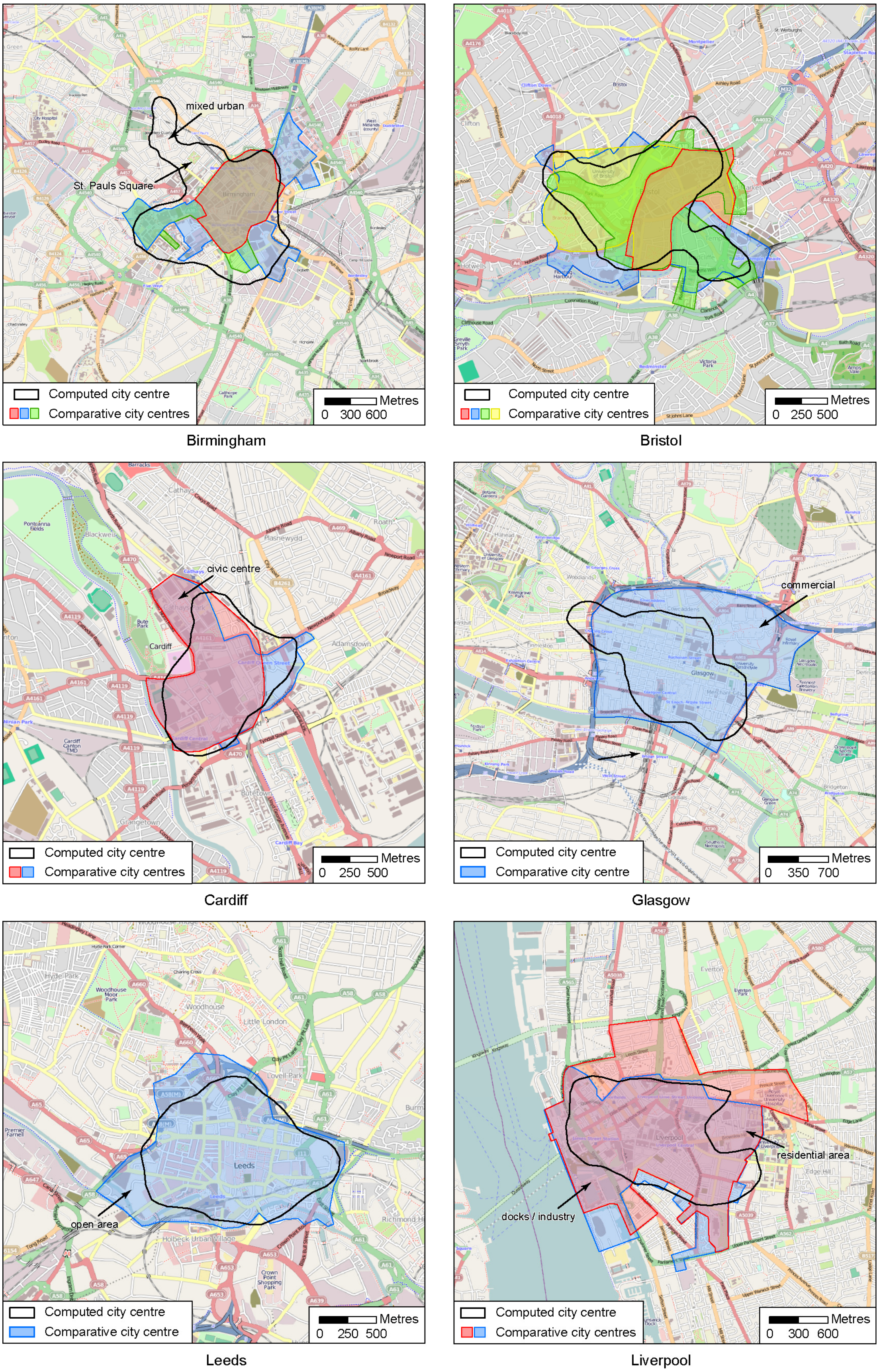


Fig. 8. Delineated city centres. Background mapping © OpenStreetMap contributors, CC-BY-SA.

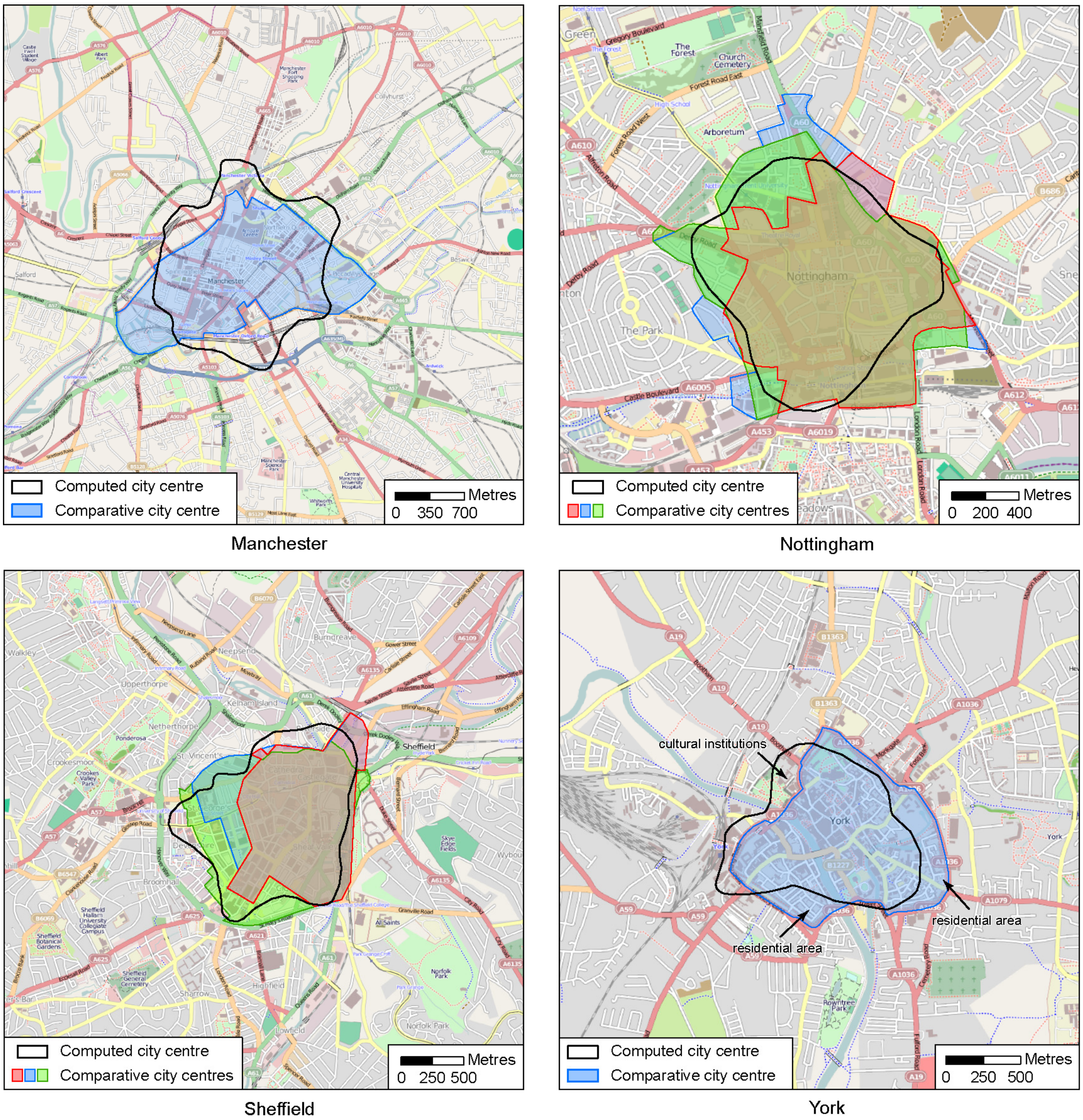


Fig. 9. Delineated city centres continued. Background mapping © OpenStreetMap contributors, CC-BY-SA.

City	Comparative – Intersection			Comparative – Union			Flickr – 80% Volume Contour		
	Precision	Recall	F_1-score	Precision	Recall	F_1-score	Precision	Recall	F_1-score
Birmingham	0.34	0.97	0.50	0.61	0.77	0.68	0.62	0.90	0.73
Bristol	0.30	0.86	0.45	0.94	0.62	0.75			
Cardiff	0.72	0.78	0.75	0.89	0.77	0.82			
Glasgow	0.92	0.51	0.65	0.92	0.51	0.65	0.98	0.68	0.81
Leeds	0.97	0.71	0.82	0.97	0.71	0.82			
Liverpool	0.88	0.67	0.76	0.96	0.46	0.62	0.79	0.63	0.70
Manchester	0.65	0.77	0.71	0.65	0.77	0.71	0.90	0.87	0.88
Nottingham	0.80	0.84	0.82	0.99	0.68	0.81			
Sheffield	0.62	0.93	0.74	0.91	0.82	0.86			
York	0.85	0.74	0.79	0.85	0.74	0.79			

Table 4. Comparison of overlap between computed and comparative city centres, and between computed and Flickr delineated city centres.

Owing to the uncertainties inherent to the comparative representations (as discussed in Section 3.6), a discrepancy between computed and comparative city centre is not necessarily due to an error of the computational model. To assess the *plausibility* of delineated city centres, large differences between the two types of representation were additionally investigated by looking at the streetscape in Google Street View.

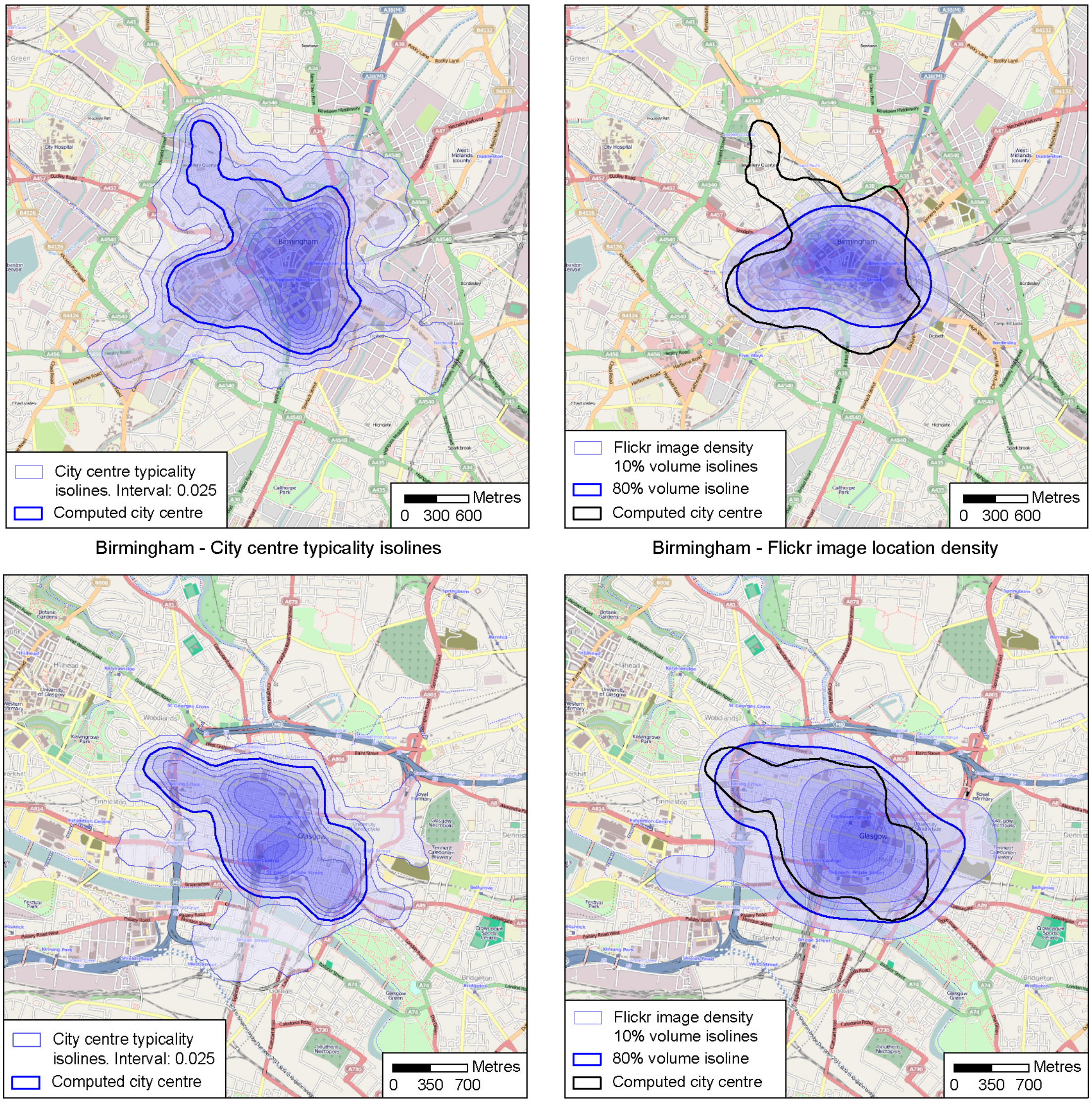
The city centres agree rather well in most cases. In Cardiff, Cathays Park was not entirely included; it hosts buildings of public administration, museums and higher education, which are arranged around a central square. The computational model omitted the area due to the high proportion of green space and the absence of other city centre functions. In Leeds, the main difference is an open area under redevelopment which was not captured as city centre by the computational model. The computed city centre in Liverpool is smaller than the comparative city centre. The main differences are residential and industrial areas not captured by the computational model. Since both comparative areas were derived from tourist maps, these areas are presumably designated as city centre because they contain sites of historic and touristic interest. Finally, York is an interesting case because the city centre is historically tightly confined by town walls. However, there are residential areas within the walls which were excluded, but an area hosting some cultural and public institutions outside of the wall was included.

Notable discrepancies occur for Birmingham, Glasgow, and Manchester. In Birmingham, the computational model delineated a protuberance that expands the city centre to the north-west. The area visually resembles a city centre up to St. Pauls Square. However, including the part beyond that square is rather questionable since it actually consists of a mix of different uses in mostly low-rise buildings.

4.2 Delineated city centres for Flickr image locations

Figures 10 and 11 show contour lines for computed city centre typicality on the left hand side, and densities of Flickr image locations on the right hand side. Glasgow and Manchester agree well with the distributions of Flickr image locations. In the quantitative comparison in Table 4 they achieve now high F_1-scores of 0.81 and 0.88, respectively. The agreement is better than with the comparative city centres, where only F_1-scores of 0.65 and 0.71 were achieved.

Birmingham’s city centre defined by the image locations is quite compact. The main difference to the computed city centre is again the protuberance to the north-west. For Liverpool, the image locations suggest that the city centre extends further to the west and includes the dockland.



Birmingham - City centre typicality isolines

Birmingham - Flickr image location density

Glasgow - City centre typicality isolines

Glasgow - Flickr image location density

Fig. 10. Computed city centre typicality (left) and Flickr image location densities (right) in Birmingham and Glasgow. Background mapping © OpenStreetMap contributors, CC-BY-SA.

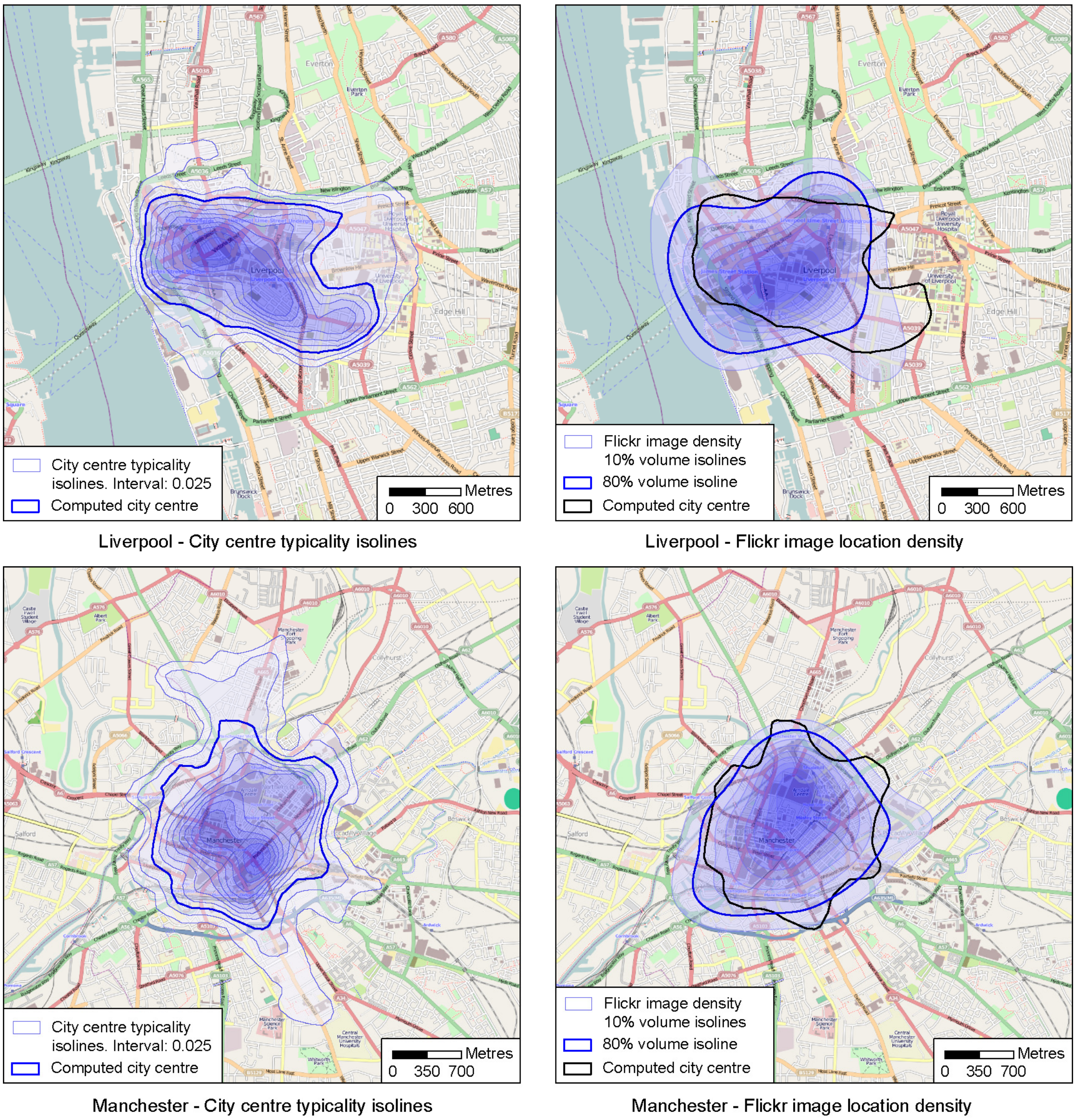


Fig. 11. Computed city centre typicality (left) and Flickr image location densities (right) in Liverpool and Manchester. Background mapping © OpenStreetMap contributors, CC-BY-SA.

4.3 Empirical city centre typicality for panorama sites

Figure 12 presents the empirical city centre typicality as it was judged by the participants based on the panoramic images. The sites were categorised into different types of environment as judged by the authors in Figure 12. Numbers in brackets indicate site numbers relating to the site locations provided in the electronic supplementary material.

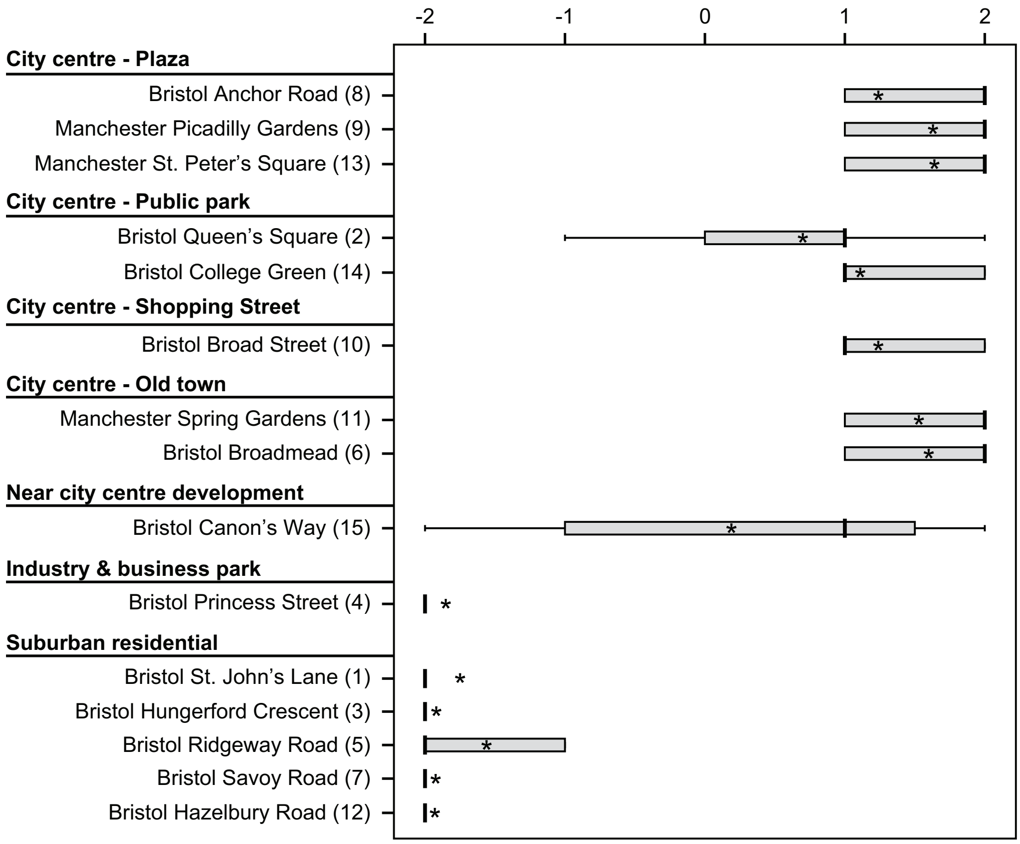


Fig. 12. Empirical city centre typicality for panoramic image sites. The box plots indicate mean (*), median (thick line), and 1st and 3rd quartile (width of boxes). Whiskers include approximately 95% of the responses.

The respondents' judgment of residential and industrial situations is clearer than that of city centre situations. There is considerable variation of perceived typicality between the different city centre categories. Evaluating the respondents' comments on their judgments, it seems that open space (in particular green space) and low rise buildings (i.e., only two or three storeys high) have a strong negative influence on perceived city centre typicality. Two sites were judged rather ambiguously: Bristol Queen's Square, which is within the city centre, but features some green space, two storey buildings and no visible shops or business; and Bristol Canon's Way, which is a new near city centre development featuring business, leisure and tourist attractions.

The computed city centre typicality was subsequently compared to the empirical values obtained for the panorama locations. Figure 13 shows a scatter plot of the empirical values and a linear least squares regression line. The squared Pearson correlation coefficient of the regression is r^2 = 0.916. Since the same set of respondents were used to elicit the knowledge for the computational model as well as for the empirical judgment of panorama sites, this cross-comparison cannot be seen as an independent validation of the computational results. Nevertheless, it shows how consistent the respondents are in their verbal descriptions of city centre functions and their visual judgment of exemplars. Furthermore, the strong correlation seems to indicate that the key functions of a city centre have been picked up by the computational model.

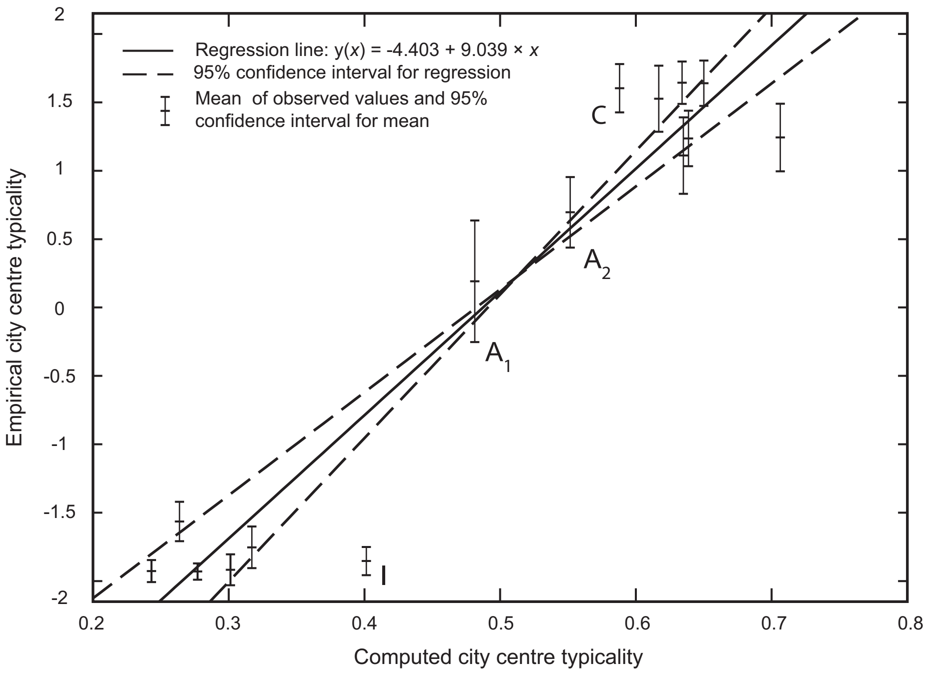


Fig. 13. Relation between empirical and computed city centre typicality.

In Figure 13 there is a cluster at very low empirical city centre typicality and one at high typicality. These clusters correspond to the selection of test sites, which were chosen to be prototypical of non-city centre and city centre situations. There is also notable agreement for the two sites Canon's Way and College Green (marked A_1 and A_2 in Figures 13 and 14), which were considered to be less clearly definable as being within or outside the city centre. Within each cluster, the variability of computed typicality is larger than the one of empirical typicality. Many of these discrepancies can be explained through the fact that the participants' judgment was restricted to those clues that were visible in the panorama, while the algorithm had information about the larger surrounding area.

Figure 14 shows the spatial distribution of city centre typicality values in Bristol. The site marked as *I* in Figures 13 and 14 is an industrial site and was thus judged as very untypical for a city centre by the respondents. But the proximity of the city centre and a high street with shopping facilities (blue stretch to the west of site *I*) leads to an increased computed city centre typicality. The site marked as *C* shows Broadmead, Bristol's city centre shopping district, and was judged as being very typical for a city centre. However, due to the remoteness to landmark buildings (which are concentrated in the cluster south-west of *C*) and other functions than shopping, the algorithm assessed the site as being less typical for a city centre.

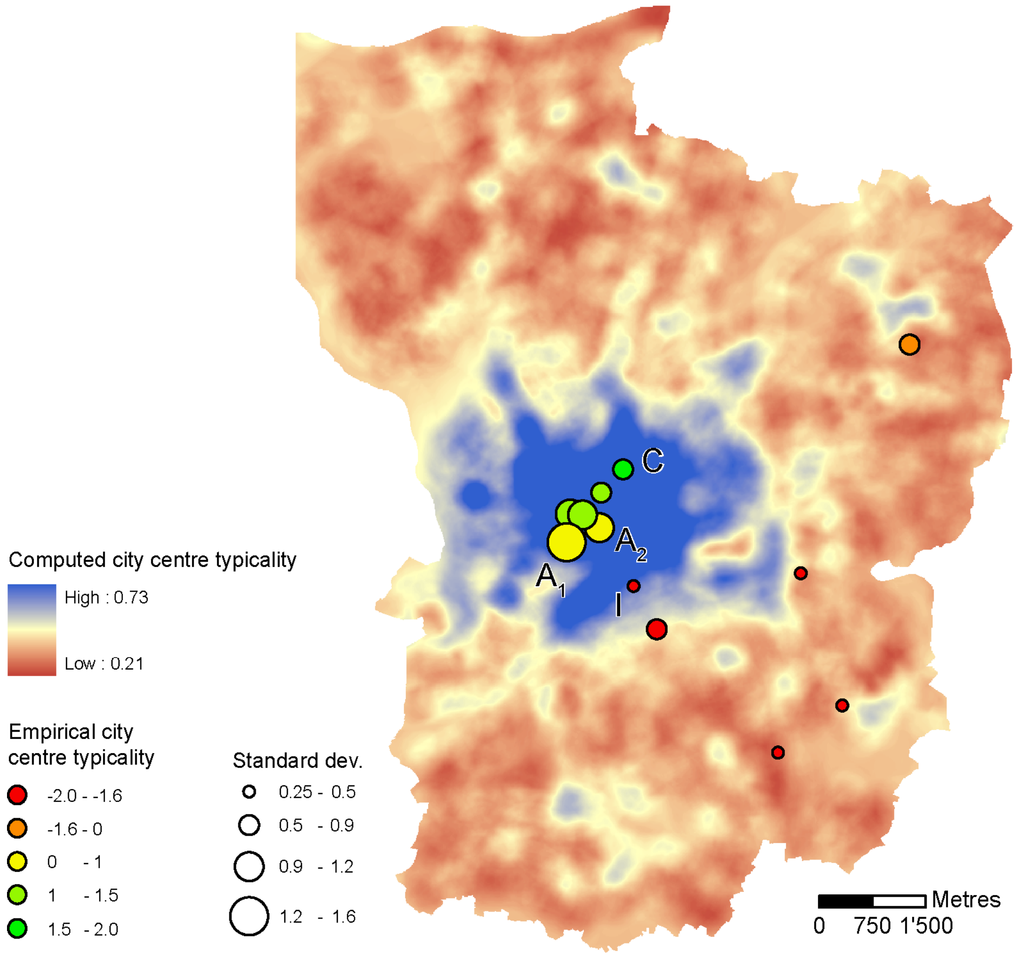


Fig. 14. Comparison of spatial distributions of city centre typicality values in Bristol.

5 Discussion

This paper argued that modelling the underlying conceptual structure is critical to enable automatic recognition of higher order phenomena from topographic databases. Conceptualisations are often hidden and tools have to be developed to render them explicit, i.e. to specify and clarify involved concepts and their logical structure (Smith & Mark, 2001). Smith and Mark (2001) and Agarwal (2004) conducted participant experiments to elicit conceptualisations for generic geographic concepts. Thomson (2009) used a questionnaire to find out how people relate land use to landscape character. The study presented in this paper is similar, but aimed at acquiring a detailed conceptual model for a single geographic concept to allow its recognition from topographic databases. A main contribution of the research is thus a top-down approach to model generalisation that employs participant experiments to obtain structural knowledge, which is subsequently used to drive the pattern recognition process. As a second contribution we demonstrated the utility of taking a functional perspective to map generalisation, which is often merely seen as a visual optimisation process.

Two alternative methods to delineate city centres from topographic maps were proposed in the literature. Boffet's (2001) experiments for defining city centre districts employed built density and building size. Heinzle and Anders (2007) proposed to use a combination of street network patterns for locating city centres, such as ring roads and star road patterns. Road patterns are highly individual to the individual history and geographical setting of each city. Also, our preliminary experiments showed that built density and building size alone are insufficient predictors, as industrial and commercial districts often have similar morphology to city centres with respect to these properties.

In the remainder of this section, the research questions posed in Section 1 shall be revisited. The research developed a general methodology of acquiring and modelling knowledge of vaguely defined geographic phenomena for model generalisation, and conducted the process specifically for British city centres.

How can empirical knowledge be formalised to delineate higher order phenomena from topographic databases?

Three tasks were presented to elicit empirical knowledge from participants. The first task asked for uninfluenced associations of city centre qualities. The second task provided lists of features as stimuli. The last task consisted of panoramic image locations and asked to rate them and reason about the clues used. Comparison of the results produced by the first task to the set of facilities named in the second task reveals some differences, which demonstrate that the type of stimulus used is critical. For example, restaurants were the second most named typical facility, but received a moderately high typicality, whereas theatres were less frequently named, but received a high typicality. In the latter case we assume that participants omit features that are rarely used, but nevertheless are seen as important defining elements (such as theatres).

The rich information produced by the questionnaire was thus analysed in a qualitative process to distil salient patterns (Ritchie & Lewis, 2003). The qualitative approach taken in the analysis, however, involved making some deliberate decisions when formulating a computational model for city centre typicality. Setting weights of individual typicality surfaces in Table 2 required careful consideration of questionnaire results, but there is some vagueness involved which might influence the results. Similarly, while data-driven methods for bandwidth selection were used, the influence radius of landmark-like features was justified by domain knowledge.

Most critical is, however, the choice of the typicality threshold, as the delineated city centre is very sensitive to this threshold. This can be seen in Figures 10 and 11, where small changes in the threshold value lead to a much better agreement (in the case of Birmingham) or worse agreement (Glasgow and Manchester) with Flickr representations. We are therefore investigating methods for setting the threshold individually for each city. For example, suburban residential and industrial areas could be used as a mask to define the approximate extent of a city centre.

What are the defining elements of a city centre?

The findings presented in Figure 4 and Table 1 generally confirm characterisations of settlement cores found in the literature (cf. Murphy & Vance, 1954; Thurstain-Goodwin & Unwin, 2000). It has to be noted that conceptualisations generally are variable among different cultures (Straumann, 2010), and urban structures are no exception (Steiniger et al., 2008). Hence, the model derived in this research is valid for British city centres only. However, the proposed methodology could also be applied to cities elsewhere and, with modification, also to the extraction of other vaguely defined geographic phenomena.

How can the produced regions be evaluated?

As field surveys as suggested by Montello (2003) are costly to conduct on a large scale, three alternative methods were used in combination to assess the plausibility of the produced regions. The vagueness of the phenomenon city centre is evidenced in the large variation between representations from different sources (Figures 8 and 9); examples are Birmingham, where there is a significant difference between individual comparative representations, and Glasgow, where the comparative representation and the Flickr representation differ considerably. To deal with this fuzzyness, 'core' regions, i.e. the intersection of comparative regions, and 'boundary' regions, the union of comparative regions, were used for quantitative comparison. Similarly, graded visualisations of computed city centre and Flickr representations in Figures 10 and 11 allow to visually comparing the internal structure of city centres.

However, there are potential biases in both methods. Comparative representations are produced by a single (or few) person(s) and are hence authoritative. Flickr representations can be systematically biased both in terms of contributing people as well as of spatial coverage (Hollenstein & Purves, 2010). For example, locations of scenic prominence are likely to be overrepresented in Flickr. A shortcoming of the method based on panoramic images is that people judge without knowledge about spatial context beyond what is visible in the images.

Keeping the above mentioned considerations in mind, our approach seems to produce city centres that conform well to the representations derived from alternative sources. F_1-scores between 0.7 and 0.8 were achieved for all cities except for Birmingham. In Birmingham, the stretch to the north-west of the city centre included by the computational model seems to be wrong. Such mixed, commercially highly active, urban areas are often hard to distinguish from 'true' typical city centre areas based on topographical information only.

Finally, it should be noted that city centres can be bounded crisply at physical discontinuities, such as city walls (see York), water courses, or major roads. While our model currently does not take account such barriers, they could be included by modifying city centre typicality in raster cells covering barriers, making them harder to cross.

6 Conclusions and outlook

Representing the world as it is conceptualised by people is of great importance in many situations when interacting with GIS (Egenhofer & Mark, 1993; Montello et al., 2003; Hollenstein & Purves, 2010). This study presented a methodology to capture conceptualisations of vaguely defined geographic phenomena and use this knowledge to drive the cartographic pattern recognition process. The concepts that are thus extracted relate to high level semantics and provide an added value to the traditional topographic data of National Mapping Agencies and other data providers. The discussed approach aids them to adapt their data for applications such as map generalisation, integration of datasets, urban planning, and geographic search. Also, since the type of data used in our approach is widely available, the approach has the potential to be applicable worldwide.

We see three main extensions of the proposed approach in future research. Firstly, the weights were determined through analysis of the questionnaire in our experiments. Previous research (Bromley et al., 2003; Hubbard, 2002; Tallon & Bromley, 2004) revealed dependencies of individual city centre use and perception from social group and age. It could thus make sense to calibrate the city centre model to different user groups in order to better represent their view of a city centre. Secondly, the same experiments should be carried out for cities in other countries in order to find out what differences there are in the conceptualisation of city centres between different regions and cultures. Thirdly, while we represented the city centre as an area, it could also be represented as a point, depending on scale (or better: map purpose). This location would be the cognitively most representative point within the city centre (the 'cognitive centre of gravity'). It would be interesting to investigate whether that point would coincide with the location of highest city centre typicality value, the centroid of the area, or the location of a landmark concept such as the town hall.

Acknowledgements

The research reported in this paper is part of the PhD project of the first author. The authors appreciate the support of the Ordnance Survey of Great Britain for funding part of the research and for provision of data. Many thanks to Ross Purves for assistance in designing the questionnaire and improving the paper. We would like to thank all the people who spent time on distributing and completing the questionnaire.

References

Agarwal, P. (2004). Contested Nature of Place: Knowledge Mapping for Resolving Ontological Distinctions Between Geographical Concepts. In M. J. Egenhofer, C, Freksa, & H. J. Miller (Eds.), *Geographic Information Science, Third International Conference, GIScience 2004* (pp.1-21). Berlin: Springer.

Barr, S. L., Barnsley, M. J., Steel, A. (2004). On the separability of urban land-use categories in fine spatial scale land-cover data using structural pattern recognition. *Environment and Planning B*, 31(3), pages 397–418.

Bromley, R. D. F., Tallon, A. R., & Thomas, C. J. (2003). Disaggregating the space-time layers of city-centre activities and their users. *Environment and Planning A*, *35*(10), 1831–1851.

Boffett, A. (2001). Méthode de création d'information multi-niveaux pour la généralisation cartographique de l'urbain. (Unpublished doctoral dissertation). Marne-la-Vallée, Université de Marne-la-Vallée.

Borgers, A. & Timmermans, H. J. P. (1986). City Centre Entry Points, Store Location Patterns and Pedestrian Route Choice Behaviour: A Microlevel Simulation Model. *Socio-Economic Planning Sciences, 20*(1), 25–31.

Chaudhry, O. Z., & Mackaness, W. A. (2008a). Automatic identification of urban settlement boundaries for multiple representation databases. *Computers, Environment and Urban Systems*, *32*(2), 95–109.

Chaudhry, O. Z., & Mackaness, W. A. (2008b). Creating Mountains out of Mole Hills: Automatic Identification of Hills and Ranges Using Morphometric Analysis. *Transactions in GIS, 12*(5), 567–589.

Chaudhry, O. Z., Mackaness, W. A., & Regnauld, N. (2009). A functional perspective on map generalisation. *Computers, Environment and Urban Systems, 33*(5), 349–362.

Couclelis, H. (1996). Towards an operational typology of geographic entities with ill-defined boundaries. In P. Burrough & A. Frank (Eds.), *Geographic Objects With Indeterminate Boundaries* (pp. 45–55). London: Taylor & Francis.

Davies, C., Holt, I., Green, J., Harding, J., & Diamond L. (2009). User Needs and Implications for Modelling Vague Named Places. *Spatial Cognition & Computation, 9*(3), 174–194.

Davies, C., Wood, L., & Fountain, L. (2005). User-centred GI: hearing the voice of the customer. *Annual conference of the Association for Geographic Information AGI'05, People, Places and Partnerships*, November 8–10, London.

Duong, T. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software, 21*(7).

Egenhofer, M. J., & Mark, D. M. (1995). Naïve Geography. In A. U. Frank & W. Kuhn (Eds.), *Spatial Information Theory. A Theoretical Basis for GIS. Proceedings International Conference COSIT '95* (pp. 21–23). Berlin: Springer.

Hart, G., & Greenwood, J. (2003). A Component Based Approach to Geo-Ontologies and Geodata Modelling to Enable Data Sharing. In Gould, M., Laurini, R., & Coulondre, S. (Eds.), *AGILE 2003, 6th AGILE Conference on Geographic Information Science* (pp. 197–206). Lausanne: Presses polytechniques et universitaires romandes.

Heinzle, F., & Anders, K.-H. (2007). Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks. In W. A. Mackaness, A. Ruas & L. T. Sarjakoski (Eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications* (pp. 233–253). Amsterdam: Elsevier.

Heinzle, F., Kopczynski, M., & Sester, M. (2003). Spatial data interpretation for the intelligent access to spatial information in the internet. In *Proceedings of the Twenty-first International Cartographic Association Conference*, Durban, South Africa.

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, No 1 (2010), 21–48.

Hubbard, P. (2002). Screen-shifting: consumption, 'riskless risks' and the changing geographies of cinema. *Environment and Planning A*, *34*(7), 1239–1258.

Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A Brief Survey of Bandwidth Selection for Density Estimation. *Journal of the American Statistical Association*, 91(433), 401–407.

Joubran, J. & Gabay, Y. A (2000). Method for Construction of 2D Hull for Generalized Cartographic Representation. In D. Fritsch & M. Molenaar (Eds.), *19th Congress of the International Society for Photogrammetry and Remote Sensing.* International Archives of Photogrammetry and Remote Sensing. IAPRS Vol. XXXIII, Part B4 (pp. 417–424).

Lloyd, C. D. (2007). *Local Models for Spatial Analysis*. Boca Raton: Taylor & Francis Group.

Lowe, M. (2005). The regional shopping centre in the inner city: a study of retail-led urban regeneration, *Urban Studies, 42*(3), 449–470.

Lüscher, P., Weibel, R., & Burghardt, D. (2009). Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. *Computers, Environment and Urban Systems*, *33*(5), 363–374.

Lüscher, P., Weibel, R., & Mackaness, W. A. (2008). Where is the Terraced House? On The Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases. In A. Ruas & C. Gold (Eds.), *Headway in Spatial Data Handling. Proceedings of the 13th International Symposium on Spatial Data Handling* (pp. 449–466). Berlin: Springer.

Mackaness, W. A. (2006). Automated Cartography in a Bush of Ghosts. *Cartography and Geographic Information Science, 33*(4), 245–256.

Mallenby, D. (2007). Handling Vagueness in Ontologies of Geographical Information. (Unpublished doctoral dissertation). Leeds: University of Leeds.

Mennis, J. L., Peuquet, D. J., & Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science, 14*(6), 501–520.

Millward, H. (2004). A Vector-GIS Extension for Generalization of Binary Polygon Patterns. *Cartographica*, *39*(4), 55–64.

Montello, D. R. (2003). Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of Geographic Information Science* (pp. 173–189). London: Taylor & Francis.

Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's Downtown?: Behavioral Methods for Determining Referents of Vague Spatial Queries. *Spatial Cognition & Computation*, *3*(2), 185–204.

Murphy, R. E., & Vance, J. E. Jr. (1954). Delimiting the CBD. *Economic Geography, 30*(3), 189–222.

Nyerges, T. L. (1991). Representing geographical meaning. In B. P. Buttenfield & R. B. McMaster (Eds.), *Map Generalization. Making Rules for Knowledge Representation* (pp. 59–85). Essex: Longman Scientific & Technical.

Office for National Statistics. (2010). Population estimates for UK, England and Wales, Scotland and Northern Ireland - current datasets. Mid Year Population Estimates 2009. http://www.statistics.gov.uk/statbase/Product.asp?vlnk=15106. Accessed 19.07.2010.

Purves, R. S., Clough, P., Jones, C. B., Arampatzis, A., Bucher, B., Finch, D., Fu, G., Joho, H., Syed, A. K., Vaid, S., & Yang, B. (2007). The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International Journal of Geographical Information Science, 21*(7), 717–745.

Regnauld, N. (2001). Contextual Building Typification in Automated Map Generalization. *Algorithmica, 30*(2), 312–333.

Ritchie, J., & Lewis, J. (Eds.) (2003). *Qualitative research practice: a guide for social science students and researchers.* London: SAGE Publications Ltd.

Rosch, E. (1978). Principles of Categorization. In Rosch, E. & Lloyd, B. B. (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Smith, B., & Mark, D. M. (2001). Geographical categories: an ontological investigation. *International Journal of Geographical Information Science, 15*(7), 591–612.

Steiniger, S., Lange, T., Burghardt, D. & Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS, 12*(1), 31–59.

Straumann, R. K. (2010). Extraction and characterisation of landforms from digital elevation models: Fiat parsing the elevation field. (Unpublished doctoral dissertation). University of Zurich, Switzerland.

Tallon, A. R., & Bromley, R. D. F. (2004). Exploring the attractions of city centre living: evidence and policy implications in British cities. *Geoforum*, *35*(6), 771–787.

Thomson, M. K. (2009). Dwelling *on* Ontology – Semantic Reasoning over Topographic Maps. (Unpublished doctoral dissertation). London, University College London.

Thurstain-Goodwin, M., & Unwin, D. (2000). Defining and Delineating the Central Areas of Towns for Statistical Monitoring Using Continuous Surface Representations. *Transactions in GIS, 4*(4), 305–317.

Townshead, T., & Pain, R (2000). Community safety in the city centre. *Town and Country Planning, 69*(4), 120-121.

Weibel, R. (1997). Generalization of Spatial Data – Principles and Selected Algorithms. In M. van Kreveld, J. Nievergelt, Th., Roos, & P. Widmayer (Eds.), *Algorithmic Foundations of Geographic Information Systems* (pp. 99–152). Berlin: Springer.

Winter, S., Tomko, M., Elias, B., Sester, M. (2008). Landmark hierarchies in context. *Environment and Planning B, 35*(3), 381–398.

Electronic supplementary material to city centre experiment

This document contains additional information about panoramic image sites (pp. 2–4) and the full city centre questionnaire (pp. 5 ff.).

The questionnaire was originally distributed online in the form of a web site, but it was reformatted to fit on paper in this document. Page breaks in the original questionnaire are indicated through "next page". Part III of the questionnaire (assessment of panoramic image sites) contains only one exemplary site. The participants had to answer the same questions for 10 sites which where selected randomly from a total of 15 sites of the study.

Panoramic image sites

12 of the 15 sites were located in Bristol. Additionally, 3 sites located in Manchester were selected to provide a more diverse coverage of city centre situations. Figures 1 and 2 show the distributions of panoramic image sites in Bristol and Manchester, respectively. A link to each site in Google StreetView is given in Table 1. They can be used to follow the cues given to the participants in the experiment. Additionally, a document containing the panoramic images as presented to the participant is available from the author's website: http://www.geo.uzh.ch/~luescher/citycentresurvey/.

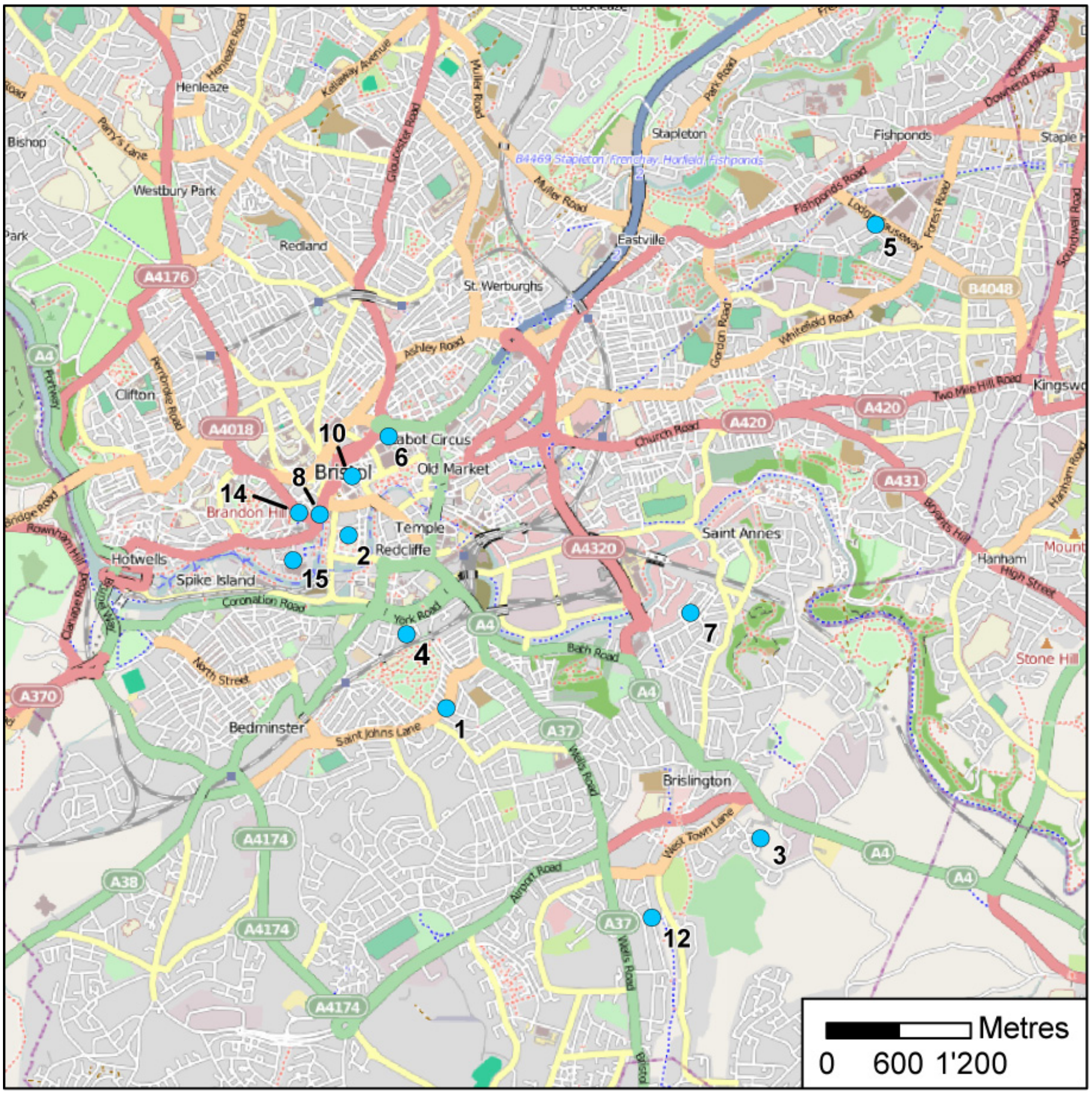


Fig. 1. Distribution of panoramic image sites in Bristol. Background mapping © OpenStreetMap contributors, CC-BY-SA.

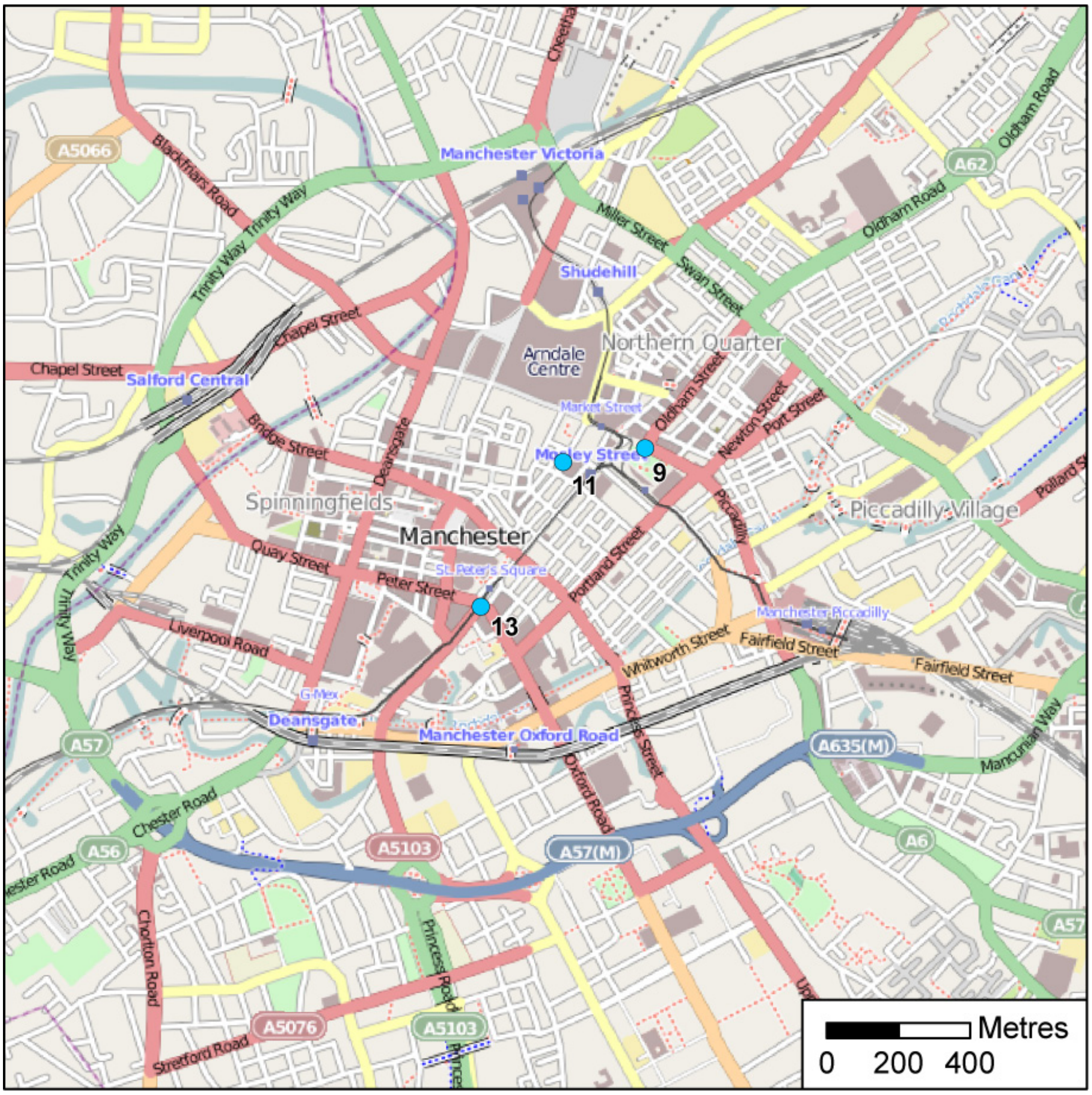


Fig. 2. Distribution of panoramic image sites in Manchester. Background mapping © OpenStreetMap contributors, CC-BY-SA.

Site Nr.	Participants who indicated location (%)	Participants who recognised location (%)	Empirical city centre typicality: Mean μ	Empirical city centre typicality: Std.dev. σ	Computed city centre typicality	Google StreetView link
1	2.74	2.74	-1.75	0.66	0.32	Click here
2	10.71	8.33	0.70	1.17	0.55	Click here
3	0.00	0.00	-1.92	0.49	0.30	Click here
4	1.30	0.00	-1.85	0.46	0.40	Click here
5	3.49	2.33	-1.56	0.68	0.26	Click here
6	7.59	7.59	1.60	0.80	0.59	Click here
7	0.00	0.00	-1.93	0.26	0.28	Click here
8	17.50	12.50	1.24	1.10	0.71	Click here
9	16.87	14.45	1.63	0.73	0.63	Click here
10	9.46	4.05	1.24	0.88	0.64	Click here
11	26.32	13.16	1.53	0.76	0.62	Click here
12	2.44	0.00	-1.93	0.26	0.24	Click here
13	38.00	38.00	1.64	0.59	0.65	Click here
14	19.15	19.15	1.11	0.96	0.64	Click here
15	0.00	0.00	0.19	1.56	0.48	Click here

Table 1. Experiment results for panoramic image sites. Click on the link in the rightmost column to access the panoramic image location in Google Maps.

Dear participant

You are invited to participate in our survey on characterisation of British city centres. It will take approximately 30 minutes to complete the questionnaire. Please note that you need to be resident within the UK to do they survey.

It is very important for us to learn your opinions. Participants that completed the questionnaire have the chance to win a gift voucher of £50 for amazon.co.uk. We are drawing three gift vouchers totaling £150.

The structure of the questionnaire is as follows:
- Part I: Participant background. [1 page]
- Part II: Text-based survey about important features of the city centre. [3 pages]
- Part III: You are shown 10 individual locations by means of a panorama taken at that location. You will be asked if the location belongs to a city centre for each location. [10 pages]

Your survey responses will be strictly confidential and data from this research will be reported only in the aggregate. Your information will be coded and will remain confidential. If you have questions at any time about the survey or the procedures, you may contact us at any time.

Thank you very much for your time and support. Please start with the survey by clicking on the **Continue** button below.

This survey is conducted by:

Patrick Lüscher
Research Associate
Department of Geography
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich (Switzerland)

Phone: +41 44 635 52 17
Email: patrick.luescher@geo.uzh.ch

next page

Questions marked with a * are required

Part I

Information about the participant's background

Your age *

-- Select --

Your sex *

- Male
- Female

What is your level of proficiency concerning the use of maps, especially conerning urban applications? This includes digital representations such as Google Maps and Open Street Map. *

- Infrequent user: I rarely look at maps.
- Casual user: I occasionally use maps for planning my activities in my leisure time.
- Student: I often use maps and spatial data because I study geography, urban planning or a related discipline.
- Professional user: I have a professional background in geography, urban planning or a related discipline.

Cultural background

These questions will help use to determine whether people coming from different places of the UK have different images of city centres.

What is your current place of residence (city or town and county): *

Postcode of your place of residence:

For how long have you been living in the UK? *

- Less than 2 years
- 2 - 5 years
- 5 - 10 years
- More than 10 years

If you have lived in other places for more than two years, please name the most recent three of these (city, county and country / one place per line):

next page

Questions marked with a * are required

Part II - Capturing important aspects of city centres

Please define, briefly, in what aspects a city centre differs from other areas of a city.
Please indicate:
1. For which types of activities do you typically go to the city centre? Which types of activities are commonly performed in city centres? *

2. What kind of services & facilities do you expect to find there (in comparison to other areas)? *

3. Is the style of the buildings, roads and squares in city centres different, and how is it different? *

4. Is there anything special that hasn't already been described?

next page

Questions marked with a * are required

Please indicate your agreement to the following statements:

	Don't know	Strongly Disagree -2	-1	1	Strongly Agree 2
A city centre is a good place to go shopping. *	◯	◯	◯	◯	◯
A city centre is a nice place to live. *	◯	◯	◯	◯	◯
Using public transport, it's easier to go to the city centre than to other places in a city. *	◯	◯	◯	◯	◯
Nightlife is most bustling within a city centre. *	◯	◯	◯	◯	◯
There are lots of places to eat out in a city centre. *	◯	◯	◯	◯	◯
Not many people live in a city centre. *	◯	◯	◯	◯	◯
You can walk around a whole city centre in a day. *	◯	◯	◯	◯	◯

next page

Questions marked with a * are required

The following lists contain certain types of concepts that are to be found commonly in urban areas. Please indicate the degree to which they are typical for a city centre.

Select '**Very typical**' if:
- You think that the concept is typically only found within a city centre.
- If you think the best location to find many of the concepts is a city centre.
- If you think the concept is very characteristic for a city centre.

Select '**Very untypical**' if you wouldn't expect such a concept in a city centre.
Select '**Can be either**' if you think the concept can be found commonly within a city centre as well as outside of it.

If you are not sure about the meaning of a concept and can't answer a question, select '**Don't know**'.

	Don't know	Very Untypical -2	1	Can be either 0	1	Very Typical 2
Department Store *	○	○	○	○	○	○
Shopping Centre *	○	○	○	○	○	○
Retail Park *	○	○	○	○	○	○

	Don't know	Very Untypical -2	1	Can be either 0	1	Very Typical 2
Nightclub *	○	○	○	○	○	○
Restaurant and Pub *	○	○	○	○	○	○
Cinema *	○	○	○	○	○	○
Theatre *	○	○	○	○	○	○
Brewery *	○	○	○	○	○	○
Leisure Centre *	○	○	○	○	○	○
Hotel or Guest House *	○	○	○	○	○	○
Office *	○	○	○	○	○	○
Factory *	○	○	○	○	○	○

	Don't know	Very Untypical -2	1	Can be either 0	1	Very Typical 2
Place of Higher Education *	○	○	○	○	○	○
Museum *	○	○	○	○	○	○
Library *	○	○	○	○	○	○
Hospital *	○	○	○	○	○	○
Law Court *	○	○	○	○	○	○

The following is a list of landmark buildings. Please specify if you think the building is usually found inside or near a city centre or if it is usually outside a city centre.

	Don't know	Never in city centre -2	1	Can be either 0	1	Always in city centre 2
Castle *	○	○	○	○	○	○
Town Hall *	○	○	○	○	○	○
Main Railway Station *	○	○	○	○	○	○
Cathedral *	○	○	○	○	○	○
Place of Worship (other than Cathedral, e.g. Church, Chapel, Mosque) *	○	○	○	○	○	○
Stadium *	○	○	○	○	○	○
Hotel or Guest House *	○	○	○	○	○	○

Below is a list of areas. Please indicate whether you think that they are commonly found within a city centre.

	Don't know	Never in city centre -2	1	Can be either 0	1	Always in city centre 2
High Street *	○	○	○	○	○	○
Business Park *	○	○	○	○	○	○
Old Town *	○	○	○	○	○	○
Public Park *	○	○	○	○	○	○

(Optional) If you think any important concept was missing in the lists above, you can enter up to four additional features below.

	Name of Concept
Concept #1	
Concept #2	
Concept #3	
Concept #4	

(Optional) If you specified any additional concepts, please rank them as well

	Very Untypical -2	1	Can be either 0	1	Very Typical 2
Concept #1	○	○	○	○	○
Concept #2	○	○	○	○	○
Concept #3	○	○	○	○	○
Concept #4	○	○	○	○	○

next page

Part III - Estimation of similarity to city centre

In the following you will be shown 10 randomly-chosen locations of British cities. For each location, you will see a 360° panorama picture taken from that location. Your last task is to judge for each location if it belongs to a city centre and indicate which hints you used for your judgement.

next page

Questions marked with a * are required

Estimation of similarity to city centre

Please have a look at the following 360° panorama. You can move around in the panorama using the **scroll bars at the bottom of the picture**.

Your task is to judge if this picture is of a city centre.

How do you estimate the similarity to a city centre of the location depicted on this page (-2 = very unlike a city centre, 2 = completely like a city centre):

	Cannot judge	-2	1	0	1	2
Similarity to city centre *	○	○	○	○	○	○

Please write briefly in one sentence or in keywords how you decided (e.g. clues such as the general setting and objects visible in the panorama) *

Do you recognize the place where this photo was taken?

- yes
- no

If you answered 'yes' to the question above, where was it taken (indicate as detailed as possible)?

next page

Thank you for participating in this survey!

Please click on the Submit button at the end of this page to complete the survey.
If you like to participate in the competition for amazon vouchers, you may leave your email address below. Tick the check box if you are interested in the scientific work that resulted from this survey.

Email Address

- [ ] I would like to hear about the results of this survey.

If you have any comment that you like to share with us, you can do so below.

submit

Part III

Appendices

Appendix A
Description of datasets

A.1 OS MasterMap®[1]

MasterMap® data produced by the British Ordnance Survey (OS) were used in all studies of this thesis. Currently, the MasterMap® product suite offers the following layers:

- *Topography Layer:* A detailed representation of the physical environment
- *Address Layer 2:* A set of postal and geographic addresses
- *Integrated Transport Network™ (ITN) Layer*: Road network and road routing information
- *Imagery Layer:* Aerial imagery of Great Britain

With the exception of the Imagery Layer, OS MasterMap® is delivered in an XML format, whereas the geometric information is encoded in GML. Usually MasterMap® layers are contained within the same XML file. Hence, a Java application was developed that extracts the relevant features from an XML file and stores them into a set of ESRI Shapefiles. Out of the available layers, only the Topography Layer and Address Layer 2 were used in this thesis. In the following, they are discussed in more detail.

A.1.1 Topography Layer

The Topography Layer is captured and updated by ground survey at the scales of 1:1,250 (urban areas), 1:2,500 (rural areas) and 1:10,000 (remote areas such as mountains),

[1] The content of this section is largely based on Ordnance Survey's product specifications available from http://www.ordnancesurvey.co.uk/

respectively. Depending on the feature type, a feature might be represented as point (for example electricity poles, or trees), line (e.g. railway tracks), or polygon. Since the interest was always on land coverage, only polygon features (XML class `TopographicArea`) were used.

A classification of features in the Topography Layer can be made by four attributes: Firstly, the Topography Layer is subdivided into nine top-level themes (Attribute `theme`), such as *Buildings*, *Land*, *Water,* or *Structures*. Another classification is given by the Attribute `descriptiveGroup`, which assigns each feature to one or more of 21 groups. `descriptiveTerm`, if present, gives further information about the feature. Finally, `make` indicates whether the nature of the represented feature is man-made or natural. Table A.1 illustrates some examples of attributions that are extracted from the Topography Layer feature catalogue linked on the MasterMap® product specification website of Ordnance Survey. An extract of the Topography Layer is shown in Figure A.1.

theme	**descriptiveGroup**	**descriptiveTerm**	**make**	**Definition**
Buildings	Building		Manmade	*“A permanent roofed construction.”*
Buildings	Glasshouse		Manmade	*“A horticultural building constructed largely of glass.”*
Land	General Surface		Manmade	*“A manmade surface area.”*
Land	General Surface	Multi Surface	Multiple	*“An area containing multiple surface types representing private residential gardens.”*
Land	General Surface		Natural	*“Areas of natural surface with no specific vegetation classification e.g. agricultural land.”*
Land	Natural Environment	Nonconiferous Trees	Natural	*“Area of trees that do not bear cones, spaced at not more than 30 m apart.”*
Water	Inland Water		Natural	*“An area of fresh water, the extent of which is captured at normal winter level.”*

Table A.1: Examples of feature definitions in OS MasterMap® Topography Layer

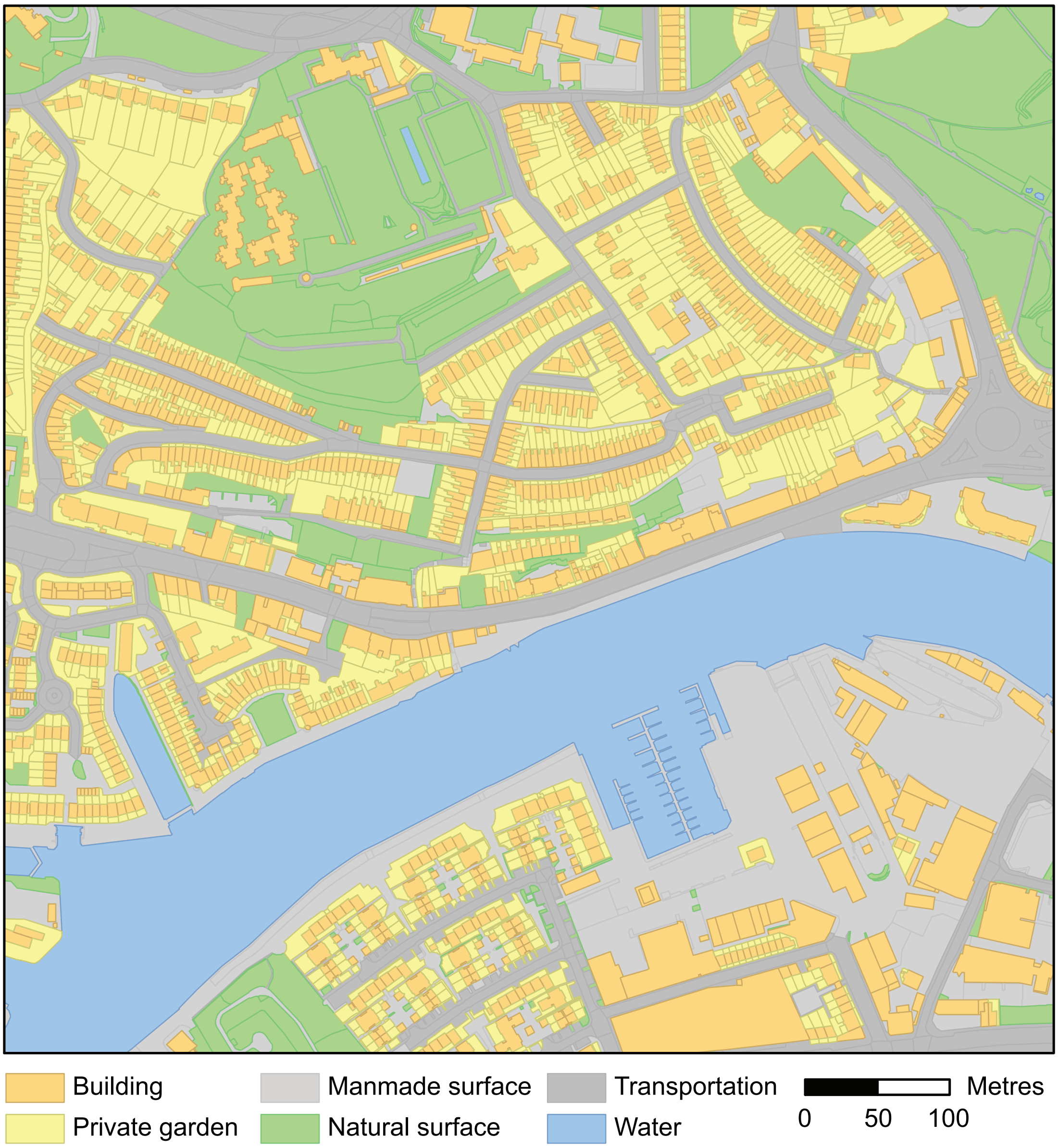


Figure A.1: Exemplary area extracted from MasterMap® Topography Layer. OS MasterMap data Ordnance Survey ©Crown Copyright. All rights reserved.

A.1.2 Address Layer 2

Address Layer 2 comprehends on the one hand postal delivery points provided by Royal Mail (*postal addresses*), and on the other hand features that do not have a Royal Mail address, but are important so that one wishes to identify them (*geographical addresses*). Examples for the latter are churches, cinemas, and car parkings. Address Layer 2 offers three different classification systems for the address points. However, one significant problem with any of these classifications is that many non-residential features are not (yet) classified. In the Bristol dataset used in the city centre experiment, 28.7% of 18,924 non-residential features are assigned to the group 'GENERAL COMMERCIAL', which can denote anything from a hospital to a cinema or restaurant. Another significant problem is the completeness of the dataset: A manual examination showed that around half of amusement establishments (night clubs, cinemas etc.) in Bristol are missing in Address Layer 2, either because they are lacking a postal address, or because the address was not classified.

Thus, in the city centre experiment that used Address Layer 2, only residential addresses were kept, while all other addresses were obtained from the Points of Interest dataset.

A.2 OS Points of Interest[2]

OS Points of Interest covers commercial and geographical addresses classified into more than 600 classes. The classes are organised into a three-level hierarchy. Table A.2 shows an extract of the classification hierarchy for the sake of illustration.

Top-level groups	**Group "Sport & Entertainment"**	**Group "Venues, stage and screen"**
Accommodation, Eating & Drinking	Entertainment support services	Cinemas
Commercial Services	Gambling	Discos
Attractions	Outdoor pursuits	Nightclubs
Sport & Entertainment	Sports complex	Social Clubs
Education & Health	Venues, stage and screen	Theatres and Concert Halls
Public Infrastructure		Conference and Exhibition Centres
Manufacturing & Production		
Retail		
Transport		

Table A.2. OS Points of Interest classification system illustrated

The dataset is provided in the form of a CSV table and can be converted into an ESRI Shapefile using ArcGIS functionality.

[2] The content of this section is largely based on Ordnance Survey's product specifications available from http://www.ordnancesurvey.co.uk/

Appendix B
Complete publication list

Listed below are all publications related to the work carried out at the Department of Geography of the University of Zurich (years 2006–2011). The publications that form this thesis are marked with an asterisk (*).

Lüscher, P., Burghardt, D., & Weibel, R. (2007). Matching road data of scales with an order of magnitude difference. *XXIII International Cartographic Conference*, Moscow, Russia, August 3–10, 2007.

* Lüscher, P., Burghardt, D., & Weibel, R. (2007). Ontology-driven Enrichment of Spatial Databases. *10th ICA Workshop on Generalisation and Multiple Representation*, Moscow, Russia, August 2–3, 2007.

Lüscher, P., Weibel, R., & Burghardt, D. (2008). Alternative options of using processing knowledge to populate ontologies for the recognition of urban concepts. *11th ICA Workshop on Generalisation and Multiple Representation*, Montpellier, France.

* Lüscher, P., Weibel, R., & Mackaness, W. (2008). Where is the Terraced House? On The Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases. In A. Ruas & C. Gold (Eds.), *Headway in Spatial Data Handling. Proceedings of the 13th International Symposium on Spatial Data Handling* (pp. 449–466). Berlin / Heidelberg: Springer-Verlag.

* Lüscher, P., Weibel, R., & Burghardt, D. (2009). Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. *Computers, Environment and Urban Systems, 33*(5), 363–374.

Lüscher, P., & Weibel, R. (2010). Semantics Matters: Cognitively Plausible Delineation of City Centres from Point of Interest Data. *Geographic Information on Demand. 13th Workshop of the ICA commission on Generalisation and Multiple Representation*, Zurich, Switzerland.

* Lüscher, P., Weibel, R. (submitted 2010). Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases. *Computers, Environment and Urban Systems*, revised manuscript submitted June 2011.

Weibel, R., Lüscher, P., Niederhuber, M., Grossmann, T., & Bleisch, S. (in press). Delivering GIScience via e-learning: The GITTA experience. To appear in D. Unwin, N. Tate, K. Foote, & D. DiBiase (Eds.), *Teaching Geographic Information Science and Technology in Higher Education*. Oxford: Wiley-Blackwell.

Appendix C

Curriculum Vitae

Patrick LÜSCHER

Date of birth: 30 May 1978
Nationality: Swiss
Place of citizenship: Muhen AG

Education

1994 – 1998	**Alte Kantonsschule Aarau, Aarau, Switzerland** Matura Typus B
2000 – 2006	**University of Zurich, Switzerland** Studies in Geography. Minors in computer science and experimental physics
2006	**Diploma in Geography (dipl. geogr.), University of Zurich, Switzerland** Diploma thesis: Matching von Strassendaten stark unterschiedlicher Massstäbe und Aufbau einer Multirepräsentationsdatenbank
2006 – 2011	**Department of Geography, University of Zurich, Switzerland** Doctoral studies in Geography. Employment as PhD student and research assistant 1 July 2006 – 31 December 2010